View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by ScholarlyCommons@Penn

University of Pennsylvania
ScholarlyCommons

Publicly Accessible Penn Dissertations

2017

Epistemic Practices In Adults And Adolescents

Sarah Emlen Metz
University of Pennsylvania, sehmetz@gmail.com

Follow this and additional works at: https://repository.upenn.edu/edissertations

Part of the Educational Psychology Commons, Epistemology Commons, and the Science and Mathematics Education Commons

Recommended Citation
Metz, Sarah Emlen, "Epistemic Practices In Adults And Adolescents" (2017). *Publicly Accessible Penn Dissertations*. 2472.
https://repository.upenn.edu/edissertations/2472

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/edissertations/2472
For more information, please contact repository@pobox.upenn.edu.

Epistemic Practices In Adults And Adolescents

Abstract
Being epistemically responsible requires an appreciation for both the power and the limitations of human knowledge, forming and adjusting one's beliefs in a way that is responsive to the right criteria. Epistemic responsibility is needed among the populace as well as the elite for a functional democracy. It is also crucial for the understanding of science. However, without clear, shared norms of how best to form, adjust, and justify beliefs, we cannot hold each other epistemically accountable. In this dissertation, I explore how adolescents and adults conceive of the best practices for forming beliefs. Chapter 1 asks what criteria for belief people take as legitimate, and how that affects their scientific beliefs. Chapter 2 examines epistemological reasoning in adolescence, focusing on appreciation for objective epistemic strategies and epistemic limitations. Chapter 3 explores adolescents' cognitive norms and capacity to engage in epistemically responsible thinking, conceived here as thinking that is actively open-minded. These questions are explored using a mixed method approach, utilizing both surveys and interviews in each chapter. Surveys allow data collection from large and diverse samples. Qualitative interviews allow us to observe reasoning and justification more directly, adding nuance and illustrating thought processes. The results reveal serious disagreement concerning what constitutes epistemic responsibility, especially with respect to what counts as a legitimate reason for belief (Chapter 1). However, there is room for optimism; the majority of adolescents demonstrated the crucial building blocks for a sophisticated epistemology (Chapter 2) and norms of actively open-minded thinking (Chapter 3). We need to leverage these existing good epistemic norms and elements of understanding, especially in the young. It is my hope that this dissertation will further this goal.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Psychology

First Advisor
Jonathan Baron

Keywords
critical thinking, epistemic responsibility, epistemological development, epistemological reasoning, nature of science

Subject Categories
Educational Psychology | Epistemology | Science and Mathematics Education

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/2472

EPISTEMIC PRACTICES IN ADULTS AND ADOLESCENTS

S. Emlen Metz

A DISSERTATION

in

Psychology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

_____

Jonathan Baron

Professor of Psychology

Graduate Group Chairperson

_____

Sara Jaffee

Professor of Psychology

Dissertation Committee

Angela Duckworth

Christopher H. Brown Distinguished

Professor of Psychology

Geoffrey Goodwin

Associate Professor of Psychology

for William James

ACKNOWLEDGMENTS

In contemplating this section, I find myself overwhelmed with gratitude. This work would not have been possible without the help of many people.

First, I would like to thank my advisor, Jonathan Baron, for his intellectual and moral support throughout the last five and a half years. When I was applying to graduate school, Jon Baron was my first choice of advisors in the world because of the clarity and rigor of thinking expressed in his groundbreaking papers and illuminating books, and because I knew he shared my goal of understanding and promoting more open, rigorous thinking. He wasn't taking students at the time, but with a lot of luck and his generosity, I became his student anyway. The more I worked with him, the more I came to appreciate his candor, wide-ranging intellectual curiosity, methodological rigor, capacity for continuous actively open-minded thinking despite strong opinions, and genuine kindness. Our conversations ranged from careful discussion of the right statistical analysis to critique of methods to politics to poetry. I will always count it an honor and a privilege to be Jon Baron's student, and I would like to thank him from the bottom of my heart for taking me on. I will strive to do you proud, and further the goals we share to the best of my ability.

I would like to thank my committee chair, Angela Duckworth, for the golden opportunity to work with a larger and more diverse population of adolescents than I had ever dreamed possible, and the freedom to pursue my own questions about intellectual virtue and epistemic responsibility. I have learned immense quantities about data collection, analysis, interviewing, and team management from working with her and her amazingly talented and hardworking team of Ducklings. Thanks especially to David Meketon for encouragement and advice; Rebecca Nyquist for her enthusiasm and help with coding; Alisa Yu for endless competence and hard work; Alyssa Matteuci for her unending good nature and helpfulness; and Annie Daeun Park for statistical advice.

I would like to thank Geoffrey Goodwin and his lab for instilling in me the habit of sniffing out confounds and other bugs in the measures and methods of every study, especially my own, and the consequent sense that no body of evidence in psychology is ever totally compelling. I continue to be inspired by his methodological and theoretical rigor, and will strive to continuously improve the rigor of my own work through critique and iteration.

Major thanks are due to Michael Weisberg and Deena Skolnick Weisberg, who supplied the opportunity for the project described in Chapter 1, and their lab, especially Jane Reznik, Kelly Kennedy, and Sabrina Elkassas for assistance in data collection and interview transcription and coding. Especial thanks to Deena for her expertise in forming and piloting survey items. Huge thanks to Michael for reigniting my fascination with psychology through reconnecting it back to the huge reservoir of deep questions in philosophy, for his support, endless enthusiasm, hard questions, and encouragement.

Thanks are also due to the friends and colleagues who have given intellectual and moral support throughout this process: thanks to Johannes Eichstaedt for reassurance and warmth in moments of self-doubt; to Justus Myers for sage advice in times of bewilderment; to Andy Gersick for wit and perspective amid exasperation; to David Yaden, Josh Baker, Corey Cusimano, Justin Landy, and Katrina Fincher for encouragement, support, and kindness, and to all of the above for their valued friendship and for keeping my intellectual gears going and sparking new ideas even when they got sticky.

The research described in Chapter 1 was supported in part by a grant from the National Science Foundation (STS-1455425 to DSW and MW). I would like to thank Kathleen Hall Jamison and the staff at the Annenberg Public Policy Center for their generous assistance with obtaining the demographically representative sample in Study 3.

The research described in Chapters 2 and 3 was supported by the Templeton Foundation. Thanks are also owed to collaborators Bill Damon, Anne Colby, Robert Franklin Borah, and Heather Malin for help in collecting data from our California schools. Thanks in addition to all the school administrators and teachers who made this research possible, and the students for their willingness to take yet another survey and their interest in sharing their thoughts and feelings with us through the interviews.

Gratitude is also due to the inestimable Christopher Teti for his infinite support, patience, and reminding me to eat; to Sebastian, Molly Wilder, Mikio Akagi, and Logan Dwyer for keeping alive the joy of the life of the mind; and to my father, Max Henrion, for showing me just how rich, valuable, and variegated such a life could be.

Last but by no means least, thanks to my mother, Kathleen Metz, without whom I would have been lost, even if I'd managed to somehow exist. I don't know whether I owe her more for her inexhaustible moral support or her always-available expertise in exactly the subjects for which I needed most help: qualitative research methods (on which she had conveniently taught a graduate level seminar for years at Berkeley), stage models of development (a literature in which she had published seminal papers), and how to put together a clear and well-structured paper. Thank you for everything, Mom. This wouldn't exist without you.

This list is far from exhaustive, but if I list everyone to whom I am grateful, you'd never reach the papers. To all my other colleagues, friends, and family who have encouraged and challenged me in the last five years: Thank you!!!!!

ABSTRACT
EPISTEMIC PRACTICES IN ADULTS AND ADOLESCENTS

S. Emlen Metz

Jonathan Baron

Being epistemically responsible requires an appreciation of both the power and the limitations of human knowledge, forming and adjusting one's beliefs in a way that is responsive to the right criteria. Epistemic responsibility is needed among the populace as well as the elite for a functional democracy. It is also crucial for the understanding of science. However, without clear, shared norms of how best to form, adjust, and justify beliefs, we cannot hold each other epistemically accountable. In this dissertation, I explore how adolescents and adults conceive of the best practices for forming beliefs. Chapter 1 asks what criteria for belief people take as legitimate, and how that affects their scientific beliefs. Chapter 2 examines epistemological reasoning in adolescence, focusing on appreciation for objective epistemic strategies and epistemic limitations. Chapter 3 explores adolescents' cognitive norms and capacity to engage in epistemically responsible thinking, conceived here as thinking that is actively open-minded. These questions are explored using a mixed method approach, utilizing both surveys and interviews in each chapter. Surveys allow data collection from large and diverse samples. Qualitative interviews allow us to observe reasoning and justification more directly, adding nuance and illustrating thought processes. The results reveal serious disagreement concerning what constitutes epistemic responsibility, especially with respect to what counts as a legitimate reason for belief (Chapter 1). However, there is room for optimism; the majority of adolescents demonstrated the crucial building blocks for a sophisticated epistemology (Chapter 2) and norms of actively open-minded thinking (Chapter 3). We need to leverage these existing good epistemic norms and elements of understanding, especially in the young. It is my hope that this dissertation will further this goal.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

CHAPTER 3

LIST OF ILLUSTRATIONS

CHAPTER 1

NON-SCIENTIFIC CRITERIA FOR BELIEF SUSTAIN COUNTER-SCIENTIFIC BELIEFS

Abstract:

Why is there widespread public disagreement about evolution when the scientific consensus is so clear? We propose a novel explanation: allegiance to different criteria for belief. In one interview study, two online surveys, and one nationally representative phone poll, we confirm that evolutionists and creationists take different justifications for belief as legitimate. Those who believe in evolution emphasize openness to empirical evidence and scientific consensus (i.e., publicly accessible reasons). Creationists emphasize not only the Bible and religious authority, but also knowledge of the heart (i.e., private reasons of faith or intuition). Theistic Evolutionists fall in between Creationists and Evolutionists. These criteria for belief remain predictive of views about evolution even when taking into account other related factors like religion, political affiliation, and education. Each view is thus supported by its own internally specified criteria for what constitutes a justified belief. Changing minds may thus require a conversion of epistemic norms.

Introduction

The American public has generally positive attitudes towards science. In the latest Pew poll, 79% of Americans said science has made life easier and 72% approved of government spending on scientific research (Funk & Rainie, 2015). Yet several well-established scientific theories still struggle to gain acceptance in the public at large, particularly evolution and climate change. Despite considerable scientific evidence and nearly complete consensus among scientists, these theories remain controversial among citizens and politicians. As of 2014, only 40% of Americans agreed that human-caused climate change is happening (Pew, 2014), while 50% accepted evolution (Gallup, 2014). Why does an essentially science-respecting population resist both scientific consensus and scientific evidence on these issues?

In the present attempt to understand this puzzle, we focus primarily on evolution, which is a relatively old controversy. Many of the early arguments against evolution are still used today, and

met with the same responses, giving the debate an increasing flavor of stalemate. American skeptics seem to be particularly recalcitrant. In a survey of 34 nations, the United States had one of the lowest rates of acceptance, outdone only by Turkey (Miller, Scott, & Okamoto, 2006). The latest Gallup poll showed 42% of Americans rejected evolution entirely, while 31% accepted evolution with God's guidance, and a mere 19% accepted evolution without God's help (Gallup, 2014). The controversy over evolution is thus a good starting place to understand epistemic aspects of resistance to well-confirmed scientific ideas.

The literature suggests at least four possible sources for this resistance: inadequate understanding of the theory of evolution (Shtulman & Calabi, 2008), inadequate understanding of the nature of science (Lombrozo, Thanukos, & Weisberg, 2008), inadequate analytic reasoning (Gervais, 2015), and the custom of forming beliefs aimed at cohesion with one's community and social identity rather than truth (Kahan, 2016). We will briefly describe these four, and then propose a fifth source of resistance to evidence: disagreement over which criteria for belief are legitimately justifying. Candidate criteria for belief include both those approved as legitimate by science, such as the criterion of fit with scientific evidence, and those rejected by science, such as the criterion of fit with feelings. Differences in criteria for belief may even be a mechanism for some of the known factors; inadequate understanding of the nature of science and/or inadequate analytic reasoning could lead to inadequate appreciation for the truth-tracking power of scientific criteria for belief, while a strong commitment to protecting one's ego and social connections could encourage the use of counter-scientific but identity-protective criteria for belief like "my friends would be upset if I didn't believe this."

Understanding of Evolution. Perhaps the most natural explanation for Americans' rejection of the theory of evolution is a lack of understanding. One remarkable feature of the theory of evolution by natural selection is its incredible explanatory power. It can be used to

explain an enormous array of biological phenomena, from the familiar (the position of our noses, the flight of birds, the ability of humans and dogs to communicate emotion) to the bizarre (infanticide in lions, coital cannibalism in mantises, island gigantism). However, the usual level of understanding of natural selection seems to be very low even among students who have just taken a course on the subject (Bishop & Anderson, 1990; Lawson & Worsnop, 1992). Among the general public, understanding of evolution is woefully inadequate for revealing its explanatory power (Metz, Skolnick Weisberg, & Weisberg, 2017). Even among students of biology, misconceptions of natural selection are common and robust (Gregory & Ellis, 2009; Kelemen & Rosset, 2009; Stover & Mabry, 2007). Without the mechanism of natural selection, evolutionary theory's explanatory power collapses, making the theory far less compelling.

Yet the evidence for the relation between acceptance and understanding of evolution is mixed. Shtulman & Calabi (2008) found this relation among undergraduates in a course on evolution: Increases in understanding of evolution were associated with increases in acceptance. However, several studies of high school students in biology courses have failed to find any relationship at all (Lawson & Worsnop, 1992; Demastes, Settlage, & Goode, 1995). Even experimental studies of undergraduate courses on the existing evidence and mechanisms for natural selection often find minimal effects on acceptance (Bishop & Anderson, 1990; Ingram & Nelson, 2006). However, this may be because the level of understanding of evolution remains shockingly low, even after significant instruction. It thus remains unclear to what extent understanding of the theory may influence acceptance of it.

Nature of Science. Some researchers have suggested that rejection of evolution is produced by inadequate understanding of the nature of science (NoS). For instance, many seem to think a 'theory' is little more than a guess, which if true would justify skepticism (Johnson & Peeples, 1987). School districts in Cobb County, GA, and Dover, PA, sought to put stickers on their

biology textbooks saying that evolution was a mere theory and therefore lacked the full legitimacy of a "scientific fact" (Pennock, 2004), encouraging this confusion between colloquial and scientific uses of "theory." Johnson and Peeples (1987) found a correlation of .45 between scores on a NoS questionnaire and acceptance of evolution among 1,812 undergraduates. Among American adults, Lombrozo, Weisberg, and Thanukos (2008) found a correlation of 0.40 ($p<.01$). This was as strong as the negative correlation of acceptance with religiosity. Lack of acceptance may thus be partially driven by lack of appreciation for the nature and practice of science.

Analytic Thinking. Cognitive reflection or analytic inclinations may be another factor. Gervais (2015) found that undergraduates at a large southern university were more likely to accept evolution and reject creationism if they got at least one item correct on the Cognitive Reflection Task (Frederick, 2005). Analytic thinking most likely acts in connection with understanding of evolution and science in general; analytic inclinations may encourage investigation of the evidence, leading to increased understanding of the epistemic power of science in general and/or evolution in particular.

Affiliation. Contrary to these explanations, which link acceptance to some aspect of understanding, Kahan (2016) found that general science knowledge (knowledge of basic scientific facts and methods) only weakly predicted acceptance of evolution and acceptance of climate change. Instead, acceptance and rejection tended to fall along lines of religiosity and political ideology. Christians and political conservatives were more likely to reject evolution and climate change, while the non-religious and politically liberal were more likely to accept both. Kahan takes this to suggest that individuals' positions on these controversial subjects arise more from social affiliation than epistemic justification. For most people, he suggests, truth may not be the primary goal of belief, at least in this domain. This suggestion runs sharply against traditional

philosophical conceptions of belief, wherein belief quintessentially aims at truth (e.g. McKay & Dennett, 2009; Williams, 2002).

Criteria for Belief. Any of these explanations for Americans' low acceptance of evolution may be correct; indeed, we suspect each plays a role. Here, we propose a fifth, complementary explanation: People may be arriving at different beliefs based on different sets of *criteria for belief*. We use such criteria to fix our beliefs and justify them to others. Without agreement on which criteria are legitimate, there will be no agreement on which beliefs are justified.

We suggest that disagreement about which criteria are the most effective guides to truth split along social and ideological lines. Each community's set of accepted criteria for belief could then support the community's beliefs, making them seem more epistemically justified to those within the group. For instance, consider a community that accepts both (a) the belief that humans possess souls with a truth-recognizing capacity and (b) the criterion for belief *it feels true in my heart.* The legitimacy of the criterion 'feeling true' can be explained and justified by the possession of a truth-recognizing soul. Meanwhile, *it feels true* supports the highly intuitive belief in souls. The framework of beliefs and criteria is thus not merely internally coherent but also self-reinforcing.

Which criteria for belief are legitimate is a question of norms. Like other norms, epistemic norms—what is taken as an adequate justification for a belief—may be socially learned to some extent. Few doubt the epistemic justifications accepted as legitimate within their community. For the most part, those raised to take the word of the Bible, their parents, and gut instinct as legitimate criteria may use these authorities to fix and justify their beliefs. Those in religious communities thus wind up with criteria that support creationism, which they were also taught to believe and which is fully justified on the basis of their accepted criteria for belief. Meanwhile, those raised to take scientific consensus as legitimate, with an emphasis on publicly available

evidence and replicable demonstrations, will mostly continue to use scientific authorities to select and justify their beliefs. These people would wind up with criteria that support the theory of evolution *they* were taught. I conjecture that only a small minority opt to critique and swap out the criteria with which they were raised. Skeptics lapse from their family religion. Converts yield hard science to the warmer rewards of faith. These few switch their epistemic priorities. I suspect that the changing emphases on different criteria play a role in subsequent shifts in belief.

There is evidence that criteria for belief can be changed by social or educational pressures. For instance, Ryu and Sandoval (2012) studied deliberately instituted norms of argumentation and criteria for belief in 3rd and 4th graders. Over the course of a year, the students significantly improved in their ability to construct and evaluate scientific arguments, with respect to causal coherence ($p<.05$), use of evidence ($p<.001$), and explicit justification ($p<.001$). Moreover, their evaluation of arguments improved substantially. At the end of the year, their justifications for these evaluations drew heavily on the criteria for belief emphasized by the teacher, indicating the students had appropriated these criteria. Moreover, videotapes of the class revealed that, as the year progressed, students increasingly came to apply and even enforce the criteria as normative.

Similarly, Kuhn, Zillmer, Crowell, and Zavala (2013) found that social discussion about moral and philosophical dilemmas among adolescent peers could develop norms of argumentation, increasing the frequency of directive meta-argumentation statements like "You need evidence for that claim" or "Tell us where you got that evidence." They were also more likely to use evidence themselves than a control group from the same school who did not engage in such discussions (Kuhn & Moore, 2015).

Insofar as criteria for belief do play a role in belief change, they could be an important mediator of the factors of belief previously discussed: NoS, analytic reasoning, group affiliation,

and even religiosity. Understanding the nature of science involves understanding why scientific criteria are so good at tracking truth. Many of the items in the Lombrozo et al. (2008) NoS scale gauge understanding of scientific rigor, for instance, "The same hypothesis or theory is often tested in many different ways." Those who do not understand the nature of science may nevertheless be taught to accept scientific criteria via social norms, but those who do understand it are more likely to accept scientific criteria because they recognize *why* those criteria are such good truth-trackers. The latter group's use of scientific criteria should be more robust to changes in social pressure.

The correlation between evolution acceptance and analytic reasoning (Gervais, 2015) could also rest in part upon the role of criteria for belief. Analytic thinkers are by definition less likely to trust their intuition or 'feelings.' Furthermore, analytic thinking can engender doubts about authorities whose influence is derived from mere circumstance, such as the people who happen to be one's parents or the religion in which one happens to have been raised. Analytic thinkers may thus be less likely to take the word of such authorities as epistemically compelling. Instead, they may prefer criteria for belief that can in principle be applied in the same way by anyone with sufficient expertise, such as the criteria of science.

To sum up, the criteria for belief taken as legitimate are likely to differ across individuals and communities. These differences in criteria should support some different beliefs. Moreover, NoS and religiosity could both shape which criteria for belief are taken as legitimate. Such criteria could thus plausibly be one mechanism by which these factors lead to the rejection or acceptance of controversial beliefs like evolution and climate change.

Based on these arguments, we hypothesized that (a) different religious and ideological communities would differ with respect to which criteria for belief are taken as legitimate, and (b)

the constellation of criteria seen as legitimate by individuals would predict rejection/acceptance of evolution and other scientific beliefs. To investigate these hypotheses, we conducted four studies. In Study 1 (*N*=47), we used one-on-one interviews to explore the diversity of epistemic justifications for beliefs about the origin of species and the extent to which these justification types diverged between creationists, theistic evolutionists, and evolutionists. In Study 2 (*N*=212), we conducted a survey to test whether these different groups would prefer different constellations of candidate criteria even when scientific or religious questions were not explicitly cued. Study 3 (*N*=965) recruited a demographically representative sample of participants to test (a) the correspondence of these differences with differences of religion, political ideology, education, and understanding of NoS and evolution and (b) the hypothesis that criteria for belief could be a mediator for two known correlates, religiosity and NoS. Study 4 (*N*=113) tested the predictive power of criteria for belief for scientific claims other than evolution. Together, these studies offer a robust and converging set of evidence for a strong link between criteria for belief and controversial beliefs.

Study 1: How Do People Justify their Beliefs?

As an initial foray into eliciting the sorts of justifications people give for scientific and counter-scientific beliefs, we conducted open-ended one-on-one interviews. This allowed us to collect a broad and illustrative sample of the kinds of justifications given for such beliefs, which we used in developing survey stimuli for Studies 2-4.

Method.

Sample. Four trained interviewers conducted 47 interviews about the origins of humans and animals with people recruited from the interviewers' four very different home communities in urban, suburban, and rural areas on the East Coast and Midwest. Participants were recruited by

the four interviewers from community networks, at local community centers, and in public spaces (e.g. a park and a train station). The sample was 48% female (M_{age}=47 years, *SD*=18.9). Thirty-five percent reported their religion as Catholic, 34% as other Christian, 12% as Muslim, and the remaining as spiritual or not religious.

Protocol. Interviewers were trained to avoid communicating either approval or disapproval of what the interviewees said, to minimize demand artifacts. This was crucial, as we wished to find out what kinds of justifications the interviewees themselves thought sufficient, without following cues of satisfaction or dissatisfaction from their interlocutor.

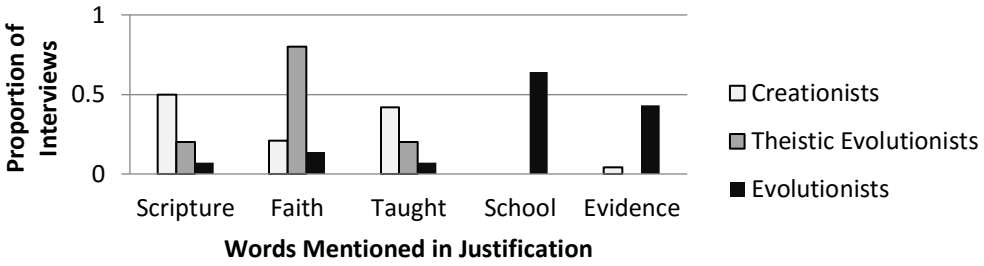
Interviewers began by asking, "How did humans come to exist?" This was followed by the questions, "How do you know?", "Is there anything that would make you more sure?", and "Is there anything that would make you less sure?" in order to elicit epistemic justifications. They then asked if the origin of other animals was the same or different. Finally, interviewers asked the interviewees to explain the theory of evolution, why scientists accepted it, and what it meant that it was a "theory." Demographics were collected with pencil and paper after the interview was completed. See Appendix for complete protocol.

Analysis. Three independent coders analyzed the transcribed interviews. The coding schema included assignment to one of three groups based on participants' accounts of the origin of humans: Creationist (humans were created by God more or less as they are today), Evolutionist (humans evolved via a natural process, with no mention of God), or Theistic Evolutionist (humans evolved via a natural process with initiation or guidance by God) (three coders: Fleiss' κ=0.86). The other codes reported below were based on the usage of particular words in response to key questions.

Results. Among the 47 interviewees, 24 (51%) were Creationists, 5 (11%) were Theistic Evolutionists, and 14 (30%) were Evolutionists. Eight expressed some ambivalence between Creationism and Evolution, including three Theistic Evolutionists, two Creationists, and three who were too ambivalent to be coded in any of the three categories. None of the Evolutionists expressed significant ambivalence. One respondent claimed humans came from aliens; he will not be further discussed.

Aside from a shared appreciation for explanatory power and the need for some form of justification, Creationists and Evolutionists gave very different kinds of epistemic justifications (see Figure 1). Creationists tended to base their justifications for their belief in the Bible, in their upbringing, or in the way they felt. Half (12/24) of Creationists mentioned Scripture (the Bible or the Quran), as opposed to 20% (1/5) of Theistic Evolutionists and only 7% (1/14) of Evolutionists. A typical Creationist justification ran, "I know that through my knowledge in the Holy Qur'an, which is the book of my belief as a Muslim." Among Creationists, 21% (5/24) mentioned *faith*, as did 80% (4/5) of Theistic Evolutionists and 14% (2/14) of Evolutionists. Of Creationists, 42% (10/24) mentioned their upbringing or education, e.g. "I was taught this growing up in a Catholic family."

Figure 1: Incidence of Words Used in Justifications for Beliefs about Human Origin



Evolutionists also commonly cited education (64%: 9/14), albeit more often by school teachers than by parents: e.g., "I think it is a combination of probably how I was brought up, [and] the fact that I've studied biology." Indeed, the two groups used different terms to describe the role of education in their beliefs. All cases of Theistic Evolutionists and Creationists speaking of education used the word "taught." In contrast, only one Evolutionist used the word "taught," the others speaking rather of "school." Creationists were more likely to emphasize parts of their education that happened outside of school, such as what they had been taught at home or at church.

Evolutionists also commonly (43%: 6/14) cited some kind of concrete evidence (fossils, DNA, or homology). For instance, one Evolutionist said,

> Certainly the archaeological record. The oldest fossil for a while was Lucy. You know, they kind of discovered these humanoid fossils – they used carbon dating to trace these back further and further in time and kind of filled in the blanks. You know again, it's not the same as plotting mathematical points, but certain use within statistical, reasonable analysis. You look at homological characteristics and how they might conceivably evolve into what we have today and kind of again, tracing these things back further and further in time, it makes sense.

In contrast, only one Creationist mentioned concrete evidence: a mention of fossils when asked to explain the theory of evolution.

As noted above, eight respondents expressed epistemic ambivalence about the question, saying they believed in both creationism and evolution, albeit in different ways and for different reasons. For instance, one subject said,

> I think there's probably two ways to look at it. I think spiritually-wise, I think that God made humans. But it looks, from an evolutionary model, that the origins of humans probably came from Africa. I think from my faith-wise… that's… through the teachings. But from the African experience… I focused on the research and journals that I've read.

As illustrated here, the justifications given by these ambivalent participants for each of their conflicting beliefs tended to fall along the same lines as those given by the more definite Creationists and Evolutionists. The inclination to accept evolution was based in school or research, while the inclination to accept creationism was based in some aspect of religion. These ambivalent participants illustrate all the more vividly the use of different justifications for different beliefs.

Discussion. Taken together, the justifications for belief given in the interviews differed systematically across the three groups, as predicted. Creationists cited feelings and upbringing as well as Scripture and faith, while many Evolutionists spoke at length about schooling and evidence. The few theistic evolutionists in our sample spoke of a more inclusive range of justifications, mentioning both faith and schooling. They seemed to have settled on theistic evolution because they felt the pull of both religious and scientific criteria, and theistic evolution was responsive to both sets of considerations.

Giving a particular justification for a claim implies that the speaker takes that justification as a legitimate criterion for belief. If it were not taken as a legitimate criterion for belief, it could not serve as a justification but at best as a causal explanation. The fact that those subscribing to

different beliefs about the origin of species gave different justifications for those beliefs suggests that they take different sorts of criteria for belief as compelling.

However, respondents may have simply brought to bear whichever criteria justified their claims. Scripture, faith, and the feeling of the heart could not reasonably be used to justify a belief in evolution, while evidence could not reasonably be used to justify a belief in creationism (although some have attempted this, e.g. the website "Answers in Genesis"). The use of criteria which do support one's claims is rational, and does not necessarily imply that those with different beliefs accept different criteria.

Respondents may even have taken the demand for justification—*How do you know?*—to be a request for the *causes* of their beliefs, rather than rational justifications for them. One woman, asked how she knew, answered, "I *don't* know; it's just what I was taught," implying that she recognized being "taught" as the explanatory *cause* of her belief but did not take it as a legitimate epistemic *justification*. The interviews thus demonstrated only differences in *use* of criteria for belief in this particular context, not necessarily differences in the perceived *legitimacy* of those criteria for belief.

Studies 2-4 aim to address these concerns, asking respondents to evaluate the legitimacy of candidate criteria in general, rather than in the specific context of evolution or science. These three surveys investigated whether differences in criteria for belief persisted outside the context of controversial topics. In each, participants first evaluated the legitimacy of a number of possible criteria for belief, with domain unspecified. Only then did they describe their beliefs and religious, ideological, and demographic identities. Answers to a multiple-choice question about evolution divided the samples of each study into three groups: Creationists, Evolutionists, and Theistic Evolutionists.

Study 2: What Criteria for Belief Are Taken as Legitimate?

Method.

Sample. We recruited 212 mTurk participants (56% female; M_{age}=33.7 years, SD=10.9 years), restricting the sample to the United States. Because after the first 132 had responded, only 17 were self-reported Creationists, we ran another 80 participants restricted to 14 states with higher proportions of evangelical Christians[1] in order to include more Creationists.

Materials. To investigate individual differences in which criteria for belief are taken as legitimate, participants were first instructed to check off which of 16 possible "reasons to believe something is true" they considered "acceptable" (see Appendix B). The prompt was thus completely general; while participants were evaluating these possible criteria, they had no indication that the study was about science, much less controversial subjects like evolution or climate change. They were allowed to check off as many as they wanted. The key non-scientific candidate criteria were: *The Bible says it is true, I feel it is true in my heart, My parents told me it is true, My clergy/minister/priest said it is true,* and *Belief that it is true encourages ethical behavior.* The key scientific candidate criteria were: *There is scientific consensus that it is true, There is good scientific evidence for it,* and *A scientist said it is true.* All non-scientific criteria were thus personal, and all three scientific criteria are widely used within scientific practice. Although the testimony of a single scientist does not constitute a strong justification within science, scientific testimony is nonetheless an essential part of expert scientific practice within the contemporary climate of specialization.

There were also 8 distractor criteria which are not necessarily scientific or unscientific, like *I saw it myself* and *I learned about it from the media.* Next, they were given the same list of 16 candidate criteria, but asked to check off only those they considered "excellent reasons for

[1] Alabama, Arkansas, Georgia, Kentucky, Louisiana, Mississippi, North Dakota, Oklahoma, South Carolina, South Dakota, Tennessee, Texas, Utah, and West Virginia.

belief." This time, they were not permitted to select more than three. This set of candidate criteria was designed to allow both criterion-level analysis and analysis on composite measures of Emphasis on Scientific Criteria and Emphasis on Non-Scientific Criteria, each constructed by summing the number of criteria in that category rated as "good" with those rated as "excellent."

On the next page, respondents answered demographic questions including religiosity (operationalized by response to the question, "How important is your religious belief to you?" on a 6-point scale). Finally, they responded to a multiple choice Origin of Species question: "Which of the following statements comes closest to your views about the origin and development of human beings?" Possible responses to this question were "Human beings developed over millions of years from other forms of life, and God had no part in this process" (Evolution), "Human beings developed over millions of years from other forms of life, and God guided this process" (Theistic Evolution), "God created human beings pretty much in their present form at one time" (Creationism), or "None of these describe my view," with a text box for them to explain their view in their own terms. This question was designed to allow both categorical analysis and quantitative analysis of acceptance of evolution, with those who chose the Theistic Evolution answer coded as intermediate between pure Creationists and pure Evolutionists.

Results. The sample yielded 36 Creationists, 60 Theistic Evolutionists, 112 Evolutionists, and 4 Others, who were left out of all analyses. The average participant checked 5.9 of 16 candidate criteria as acceptable (SD=3.1, range 1-16).

As suggested by the interviews, Creationists, Evolutionists, and Theistic Evolutionists did indeed differ considerably in which candidate criteria they took as legitimate (see Figure 2). Indeed, the number of non-scientific criteria taken as acceptable and/or excellent very strongly predicted rejection of evolution ($r(211)=.65$, $p<.001$), while the number of scientific criteria

taken as acceptable and/or excellent strongly predicted acceptance of evolution ($r(211)=.53$, $p<.001$).

Non-Scientific Criteria. Unsurprisingly, the three groups differed most in their acceptance of the *Bible* criterion, which strongly predicted rejection of evolution ($r(208)=.64$, $p<.001$). This was significantly different across all three groups, with more Creationists (74%) than Theistic Evolutionists (27%: $\chi^2(1, N = 96) = 18.97$, $p<.001$), and fewer Evolutionists (less than 1%) than Theistic Evolutionists accepting this as a legitimate reason for belief ($\chi^2(1, N = 172) = 29.14$, $p<.001$).

Clergy was also a significant predictor of acceptance ($r(208)=-.29$, $p<.01$), with Creationists accepting their epistemic authority (20%) significantly more often than did Evolutionists (1%) ($\chi^2(1, N = 212) = 18.34$, $p<.001$), though no more than Theistic Evolutionists (23%). Creationists and Theistic Evolutionists were very similar in their acceptance of *parents* (37% of Creationists, 32% of Theistic Evolutionists) and *heart* (74% of Creationists, 58% of Theistic Evolutionists; differences *ns*). However, only 14% of Evolutionists accepted *parents*, significantly less than Creationists ($\chi^2(1, N = 148) = 8.24$, $p=.004$) or Theistic Evolutionists ($\chi^2(1, N = 172) = 7.28$, $p=.007$). Only 21% of Evolutionists accepted *heart*, also significantly less than either Creationists ($\chi^2(1, N = 148) = 32.86$, $p<.001$) or Theistic Evolutionists ($\chi^2(1, N = 170) = 23.02$, $p<.001$). Both of these patterns are also apparent when acceptance of evolution is taken as a continuous variable, with acceptance of the *heart* criterion predicting rejection of evolution ($r(208) = .43$, $p<.001$) more strongly than did the *parents* criterion ($r(208) = .22$, $p<.001$).

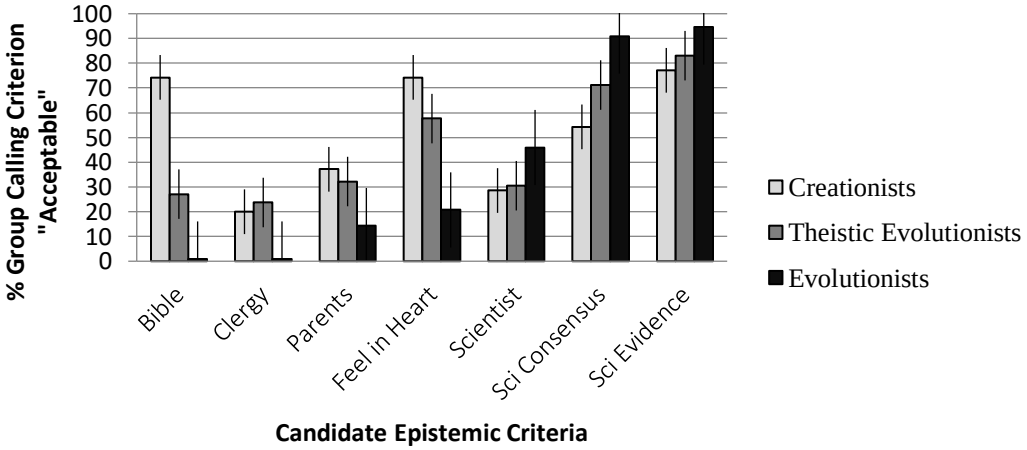
Only 12% of Evolutionists counted the *ethical* criterion as "acceptable," less than Theistic Evolutionists, who were at 27% ($\chi^2(1, N = 172) = 6.32$, $p = .01$). On this *ethical* criterion, Creationists were in the middle (17%) and not significantly different from either of the other two.

Scientific Criteria. All three groups tended to accept scientific evidence as a legitimate criterion, but it did predict acceptance of evolution (r=.22, p<.001). Although the Evolutionists were most enthusiastic (95% acceptance), most Theistic Evolutionists and Creationists accepted it, too (respectively, 83% and 77%). Nonetheless, Evolutionists were more likely than either Creationists ($\chi^2(1, N = 148) = 9.93$, p =.002) or Theistic Evolutionists ($\chi^2(1, N = 172) = 6.09$, p = .01) to accept scientific evidence as a legitimate reason for belief.

Scientific consensus was also predictive of belief (r=.35, p<.001), accepted most often by Evolutionists (91%), less often by Theistic Evolutionists (71%: $\chi^2(1, N = 172) = 11.25$, p =.001), and least often by Creationists (54%), this last group significantly less than Evolutionists ($\chi^2(1, N = 148) = 24.84$, p<.001) but not Theistic Evolutionists. *A scientist* was a weaker predictor (r=.16, p<.05), also most widely accepted by Evolutionists (46%), and less often accepted by Theistic Evolutionists (31%) or Creationists (29%) ($\chi^2(3, N = 172) = 3.95$, p = .048).

All of the group differences in *acceptability* of criteria replicated in judgments of which criteria were "Excellent," except for *clergy*, which did not reach significance.

Figure 2: Disagreement Over the Legitimacy of Criteria for Belief

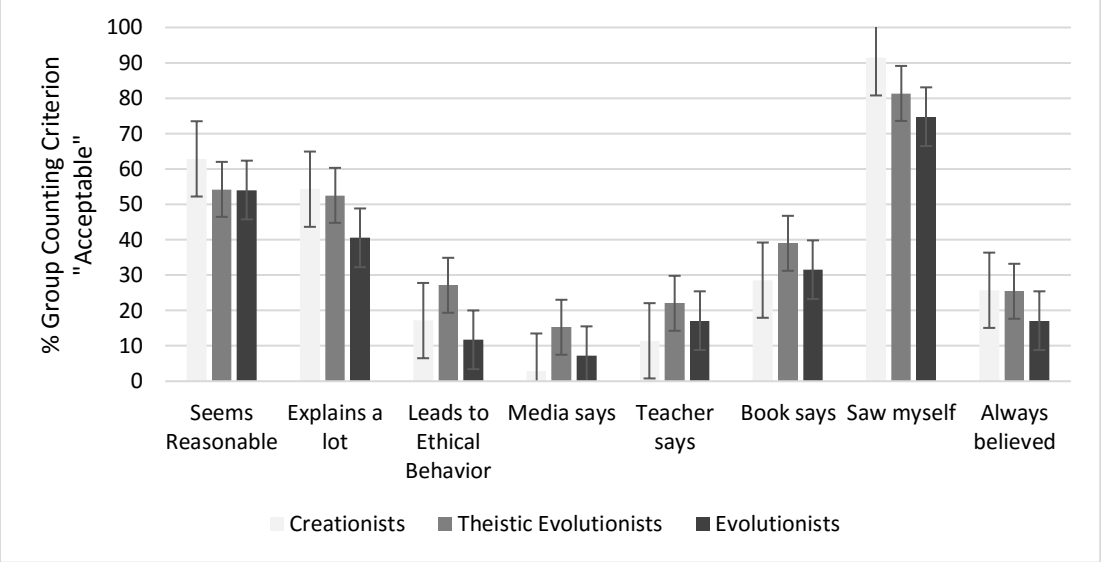


Note: Standard error bars shown.

Additional Findings. The *ethical* criterion emerged as significantly less "excellent" among Evolutionists, at only 1%, compared to 11% of Creationists ($\chi^2(1, N = 148) = 8.71$, $p = .003$) and 8% of Theistic Evolutionists (differences from other two groups not significant; differences in perceived acceptability also not significant).

However, all three groups showed equal respect for other candidate criteria, including "Explains a lot" ($\chi^2(3, N = 212) = 2.10$, *ns*) and "Seems reasonable" ($\chi^2(1, N = 212) = 3.65$, *ns*), each with about a 50% acceptance rate. There were also no significant differences between groups in acceptance of any of the distractor criteria "My teacher said so," "A book said so," "My friends saw it," "I saw it myself," or "I've always believed it" (see Figure 3).

Figure 3. Acceptance of Other Candidate Criteria



Note: Error bars show standard error.

Multiple criteria for belief contributed to predicting evolution acceptance. When acceptance of Evolution is taken as a continuous variable (1=Creationism, 2=Theistic Evolution, 3=Evolution), opinions about seven different candidate criteria significantly predict acceptance of evolution: *Bible*, *clergy*, *parents*, *heart*, *scientific evidence*, *scientific consensus*, and *scientist* (see

Table 1). The perceived legitimacy of candidate criteria also differs across demographics, especially religiosity, political ideology, and gender, but not education.

Table 1
Study 2 Pearson's Correlations of Criteria for Belief

	1. Evo	**2. Sex**	**3. Edu**	**4. Relig**	**5. Liberal**
1. Evo Acceptance	--				
2. Female	-.20**	--			
3. Education	-.02	-.12	--		
4. Religiosity	-.81**	.15*	.01	--	
5. Liberalism	.37**	.00	-.04	-.31**	--
Criteria for Belief					
6. Bible	-.64**	.15*	.02	.58**	-.28**
7. Clergy	-.29**	.04	.11	.37**	-.21**
8. Ethical	-.10	.09	-.04	.13	-.03
9. Parents	-.22**	.00	.06	.24**	-.12
10. Feel in Heart	-.43**	.24**	-.08	.44**	-.08
11. Sci Evidence	.22**	-.05	.07	-.22**	.21**
12. Sci Consensus	.35**	-.07	.08	-.29**	.16*
13. Scientist	.16*	-.14*	.03	-.09	.02
14. Explains a lot	.11	.02	.11	.10	-.02
15. Seems reasonable	-.05	-.07	.05	.00	.04
16. Always believed	-.09	.07	-.06	.12	-.06

Notes: *p<.05, **p<.01. N=208.

Discussion. These data reveal considerable disagreement in what counts as legitimate criteria for belief, substantially following the split of differences in belief about the origin of species. Even when evolution was not cued, the perceived legitimacy of candidate criteria for

belief varied considerably, indicating systematic individual differences. Furthermore, these differences in criteria for belief robustly predicted evolution acceptance. These data support the hypothesis that differences in criteria may be a source of differences in belief.

Two similarities across groups emerge. First, most individuals in every group count scientific evidence as a compelling reason for belief, although there are more science skeptics among Creationists and Theistic Evolutionists. Second, it does not seem to be the case that Creationists care any less about *explanatory power* than Evolutionists. This supports our suspicion that the problem is not any failure to appreciate the importance of explanatory power in general, but rather to appreciate the particular explanatory power of evolutionary theory. This is likely because understanding of evolution in general is fairly low; indeed, increasing understanding has been found to increase acceptance in some cases (Shtulman & Calabi, 2008).

Despite the strong effects found, it remains possible that the high correlation between criteria for belief and acceptance of evolution is an artifact of mutual correlation with one of the known correlates identified in the introduction, in particular, cognitive factors like understanding of evolution or NoS, or affiliative factors like religion or politics. Any of these factors might plausibly cause both rejection of evolution and differences in criteria for belief, leaving no causal role to the criteria. To collect further evidence regarding the relative predictive power of criteria for belief compared to already-known predictors, we recruited a larger and richer dataset, drawn from a demographically representative sample.

Study 3: Is the Criteria for Belief Effect an Artifact of Known Factors?

For Study 3, we gathered data on a greater number of possible factors of evolution rejection, so as to determine whether criteria for belief can predict rejection over and above known factors. To do this, it was important to find a representative sample of Americans, one which included the

full range of religions, political ideologies, and education levels. MTurk samples tend to skew significantly younger, more educated, and less religious than the general United States population (Berinsky, Huber, & Lenz, 2012; Simons & Chabris, 2010). For this reason, we used a professional polling firm in this study, both to attempt to replicate the findings of Study 2 in the broader population and to address the concerns above about spurious correlations due to non-representative samples.

Method.

Sample. Data were collected from 965 Americans via a professional polling firm, which used both landlines and cell phones to capture a representative sample. The questions were inserted in an omnibus survey that took about 20 minutes. Participants were allowed to skip questions. Of those contacted, 23% agreed to take the survey, and 9% of those who agreed quit partway through. The sample was 55% female, ranging in age from 18 to 96 years (M=55.5, SD=21.1). The sample was ethnically and religiously diverse: 67% Caucasian, 11% African-American, 12% Hispanic, 2% Asian, and 1.3% Native American. Ninety-two percent had at least a high school degree, 37% a bachelors' degree, and 14% had gone to graduate school. Religious affiliations were 21% Catholic, 19% Protestant, 10% "Just Christian," 5% Atheist or Agnostic, and 17% "nothing in particular." Only 4% listed a non-Christian religion; Hinduism, Buddhism, Islam, and Judaism were represented by a few participants apiece.

Materials. The questionnaire included five sections, all multiple choice or rating scales:

(1) Two items asking about two of the criteria for belief on which Creationists and Evolutionists diverged most worryingly in Study 2: "Good scientific evidence" and "Feeling it is true in your heart." Each was rated on a 5-point scale from "Excellent reason to believe something is true" to "Bad reason to believe something is true" (see Appendix).

(2) Two Likert scale items on NoS: "Once a scientific theory has been established, it is never changed," and "Scientific theories are just scientists' guesses" (both reverse scored [adapted from Lombrozo, Thanukos, & Weisberg, 2008]).

(3) Four multiple-choice items gauging understanding of evolution, with a focus on natural selection.

(4) The Origin of Species item used in Study 2.

(5) Demographics including political identity and religion.

Results. Because the survey sample was not precisely representative of the American population at large, the analyses described here were performed with population weights. However, results are robust to removing the weights.

As in Study 2, *scientific evidence* significantly predicted acceptance of evolution (F=24.71 (943), p<.001) and *heart* predicted its rejection (F=17.73 (950), p<.001). The size of these associations was quite small, r=.14 (p<.001) for *scientific evidence* and r=-.10 (p<.05) for *heart.* However, these are comparable to the associations observed with other, known factors of evolution acceptance, including NoS (r=.139, p<.001), education (r=.13, p<.001), political liberalism (r=.19, p<.001), and understanding of evolution (r=.13, p<.001). (See Table 2.)

The legitimacy accorded to scientific evidence was unrelated to the legitimacy accorded to the heart (Pearson's r=.030). Given the large sample size (N=965) and absence of ceiling or floor effects, this is strong evidence that the two criteria are genuinely orthogonal.

Table 2

Study 3 Pearson's Correlations

	1.	**2.**	**3.**	**4.**	**5.**	**6.**
1. Heart Criterion	--					
2. Sci Evidence Criterion	.030	--				
3. Nature of Sci	-.185**	.184**	--			
4. Education	-.183**	.170**	.205**	--		
5. Political Liberal	-.123**	.196**	.187**	.221**	--	
6. Understand Evolution	-.153**	.195**	.190**	.116**	.124**	--
7. Accept Evolution	-.102*	.142**	.139**	.127**	.188**	.127**

Notes: *p<.05, **p<.001.

Nonetheless, both *heart* and *scientific evidence* criteria did predict a number of other variables (see Table 2). As per our hypothesis, considering the heart a legitimate criterion predicted higher conservatism (Pearson's r=.12, p<.001), suggesting that differences in criteria for belief are tracking differences between political groups. Acceptance of the heart criterion also predicted lower education (r=-.18, p<.001), lower NoS (r=-.19, p<.001), and lower understanding of evolution (r=-.15, p<.001). Considering scientific evidence a legitimate epistemic criterion went with all the opposites: higher liberalism (r=.20, p<.001), higher education (r=.17, p<.001), higher NoS (r=.18, p<.001), and higher understanding of evolution (r=.20, p<.001).

Discussion. Study 3 replicated several previous findings in the literature as well as Study 2. Specifically, this study replicated the correlation between NoS and evolution acceptance found by Lombrozo, Thanukos, and Weisberg (2008) and the well-established correlation between religion and evolution acceptance (Sinclair, Pendarvis, & Baldwin, 1997). These patterns seem to be robust. In addition, we found evidence for an association between understanding of evolution and

acceptance of it, in line with some previous work (Shtulman & Calabi, 2008; but see Bishop & Anderson, 1990; Brem, Ranney, & Schindel, 2003).

As hypothesized earlier, one's criteria for belief could partly mediate the effects of religion and NoS on acceptance. Those whose religion and religiosity encourages them to take Scripture and clergy seriously are more likely to reject evolution on the basis of those criteria. Meanwhile, a rich understanding of the nature of science equips others to appreciate the particular predictive and truth-tracking power of scientific methods, thereby encouraging respect for scientific criteria and scientific consensus and hence increasing the probability they will accept evolution. To test this possibility, a richer measure of preferred criteria for belief is necessary.

Finally, the first three studies examined only acceptance of evolution. If criteria for belief are genuinely influencing acceptance of evolution, they should also influence and thus predict acceptance of other controversial beliefs. Study 4 was designed to address these two issues.

Study 4: Do Criteria for Belief Predict Other Scientific Beliefs?

The first three studies offered robust evidence that accepted criteria for belief are linked to acceptance or rejection of evolution, as well as demographic variables of religion and ideology. Study 4 investigated whether criteria for belief predict skepticism of other scientific claims as well, and explored the possibility that criteria for belief could be mediators of previously known factors.

Method.

Sample. We recruited 113 adult participants from MTurk, restricting participation to the United States. The resulting sample was 49.6% female, ranging in age from 21 to 64 years (M=35.2, SD=11.6). As is usual for MTurk, the sample was less religious than the American

population at large, with 36% describing themselves as not religious, 16% as Protestant, 13% as Catholic, 22% as other Christian, and less than 3% from non-Christian religious groups.

Materials. Using an online survey platform, participants answered items in the following order: (a) the legitimacy of 20 possible criteria for belief, (b) agreement with evolution and anthropogenic climate change, (c) NoS, (d) demographics, and (e) religiosity and political affiliation.

Criteria for belief. In order to capture the perceived legitimacy of various criteria for belief at a finer grain, the criteria for belief section asked participants to categorize 20 different candidate criteria as excellent, good, okay, or bad reasons for belief by dragging and dropping each from a list into one of four visual bins. Criteria in each category were aggregated to gauge acceptance of Religious Criteria (*scripture* and *clergy*) and acceptance of Scientific Criteria (*scientific evidence*, *scientific consensus*, and *scientist*). In addition, participants categorized four criteria based on affiliation (e.g. "My parents told me," "The people I love believe it is true"), two criteria based on the heart ("My heart tells me it is true," "It feels true in my heart"), three criteria based on explanatory power ("It explains a lot of the things I've seen," "It explains a whole lot of things," "It is the simplest explanation I know for some things"), one criterion based on epistemic coherence ("It fits with my other beliefs"), and several other distractor criteria (e.g. "I read it in the newspaper"). Evaluations of items within each set of criteria were quite cohesive (Scientific Criteria: α=.86; Religious Criteria: α=.78; Affiliative Criteria: α=.50; Heart Criteria: α=.89; Explanatory Power: α=.60). Items in each category were averaged to obtain overall scores for perceived legitimacy of Scientific Criteria, Religious Criteria, Heart Criteria, and Affiliative Criteria (see Appendix).

Acceptance of evolution and climate change. To find out whether accepted criteria for belief predict acceptance of climate change as well as evolution, we asked participants to rate their agreement with seven true scientific statements on a Likert agreement scale, including three items on evolution in addition to the one used in studies 2 and 3 (e.g. "Humans and chimps share a common ancestor,") two on climate change ("The climate of earth is changing significantly faster than it was 500 years ago," and "Current climate changes are at least partly caused by human activities"); and one less controversial scientific statement for comparison ("The sun is a huge ball of extremely hot gas"). Acceptance for each topic was coded as the average response to the set of questions on that topic.

Nature of science. Nine true/false questions gauging NoS were adapted from Lombrozo, Thanukos, and Weisberg (2008), e.g. "Science is only a collection of facts" (False), "Once a scientific theory has been established, it is never changed" (False), and "Scientific theories are subject to ongoing testing and revision" (True). NoS score was simply the number of items correct.

Demographics. In addition, participants were asked to specify their age, gender, political affiliation, highest level of education, and the centrality of religion to their life ("religiosity").

Results. Once again, the perceived legitimacy of scientific criteria for belief was a strong predictor of acceptance of evolution (r=.62, p<.0001). The perceived legitimacy of religious criteria was an even stronger predictor of rejecting evolution (r=-.71, p<.0001). Moreover, scientific and religious criteria both showed the same pattern not only for acceptance of climate change (scientific criteria: r=.37, p<.001; religious criteria: r=-.39, p<.001), but also for acceptance of the proposition, "The sun is a huge ball of extremely hot gas" (scientific criteria: r=.24, p<.05; religious criteria: r=-.32, p<.01 [see Table 3]).

Acceptance of the heart criteria again predicted rejection of evolution (r=-.27, p<.01), although not of climate change (r=-.09, ns). In addition, acceptance of affiliative criteria predicted rejection not only of evolution (r=-.26, p<.01), but also climate change (r=-.24, p<.05) and whether the sun is a ball of gas (r=-.22, p<.05). The criterion of *epistemic coherence*—"fit with other beliefs"—also predicted acceptance of evolution (r=.43, p<.01) and gaseous sun (r=.20, p<.05), although not climate change (r=.09, ns). Thus, a wide variety of criteria for belief predicted rejection of three unrelated scientific statements, including one that is not even considered controversial.

Table 3

Correlations of Criteria for Belief & Scientific Beliefs

Criteria for belief	Accept Evolution	Climate Change	Sun is a Ball of Gas	Nature of Science	Religiosity
Sci Criteria	.616**	.369**	.237*	.202*	-.471**
Relig Criteria	.710**	-.387**	-.323**	-.358**	.761**
Explan Power	.025	-.002	-.093	-.041	-.006
Affiliative	-.263**	-.240*	-.216*	-.303**	.237
Heart	-.272**	-.086	-.237*	-.391**	.379**
Ethical	.176	.067	.224*	.251**	-.276**
Web of Belief	.432**	.093	.198*	.268**	-.400**

Note: *p<.05; **p<.01.

The use of regression models for statistical control can offer only tentative evidence, since regression models with imperfectly measured variables may be distorted by measurement error, suppression effects, and other problems (Westfall & Yarkoni, 2016). However, consistent with our hypothesis, the predictive power of both scientific criteria and religious criteria on acceptance of evolution was sustained in a regression with NoS, religiosity, education, gender, and age (Scientific Criteria: β =.31, t=4.07, p<.001; Religious Criteria: β =-.46, t=-4.05, p<.001) (see Table 4).

Table 4

Acceptance of Scientific and Religious Criteria Predict Evolution Acceptance

Linear Regression Model	Unstandardized Coefficients		Standardized Coefficients		
	B	SE	β	*t*	*p*
(Constant)	-.973	1.021		-.953	.343
Scientific Criteria	.721	.177	**.309**	**4.069**	.000
Religious Criteria	-.751	.185	**-.464**	**-4.051**	.000
Heart Criteria	.102	.132	.060	.775	.440
Affiliative Criteria	.238	.303	.064	.785	.434
Nature of Science	.118	.089	.098	1.320	.190
Religiosity	-.070	.080	-.091	-.886	.378
Political Affiliation	.022	.068	.022	.317	.752
Level of Education	.119	.076	.106	1.574	.119
Age	-.004	.009	-.027	-.405	.686
Gender	-.186	.218	-.056	-.853	.396

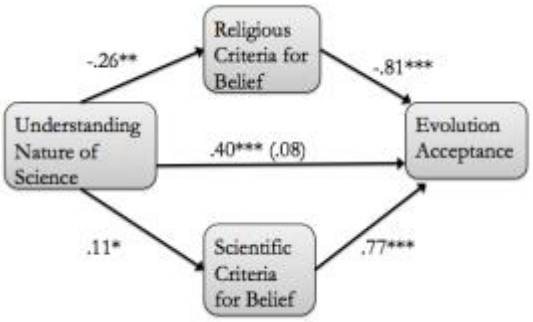
Regression analysis was used to investigate the hypothesis that criteria for belief mediate the effects of NoS and religiosity on acceptance of evolution and climate change (see Figure 3). Results again indicated that NoS was a significant predictor of evolution acceptance (b=.400, SE=.11, p<.001). NoS also predicted the perceived legitimacy of scientific criteria (b=.105, SE=.048, p=.03) and illegitimacy of religious criteria (b=-.264, SE=.067, p=.0001). Both religious criteria and scientific criteria were also significant predictors of evolution acceptance, in opposite directions: b=-.812, SE=.124, p<.0001 and b=.773, SE=.171, p<.0001. NoS was no longer a significant predictor of evolution acceptance after controlling for religious and scientific criteria as mediators (b=.010, SE=.080, *ns),* consistent with full mediation. These predictors accounted for 59% of the variance in evolution acceptance (R^2=.586). A bootstrap estimation approach with

1000 samples (Shrout & Bolger, 2002) indicates the indirect coefficient was significant for both Scientific Criteria (*b*=.081, *SE*=.041, 95% CI = .021,.188) and Religious Criteria (*b*=.215, *SE*=.061, 95% CI = .113, .362), for a total indirect effect of .296 (*SE*=.08).

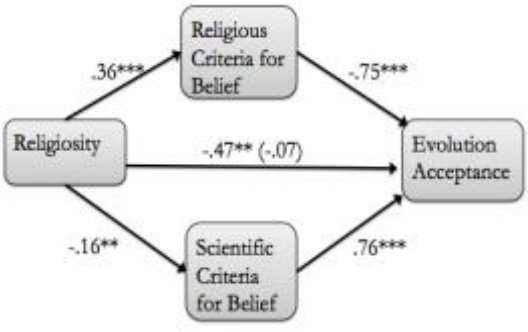
Religiosity exhibited a similar mediation pattern in its predictive power for the rejection of evolution. Although by itself religiosity is a strong predictor (*b*=-.466, *SE*=.06, *p*<.001), it becomes nonsignificant in a regression with religious and scientific criteria. Religiosity also predicted the perceived illegitimacy of scientific criteria (*b*=-.157, *SE*=.03, *p*<.0001) and legitimacy of religious criteria (*b*=.364, *SE*=.03, *p*<.0001). Religiosity was no longer a significant predictor of evolution acceptance after controlling for the two criteria variables as mediators (*b*=.072, *SE*=.075, *ns)*, consistent with full mediation. Fifty-eight percent of the variance in evolution acceptance was accounted for by these predictors (R^2=.584). The indirect coefficient (calculated again using a bootstrap estimation approach with 1000 samples) was significant for both the path via scientific criteria (*b*=-.12, *SE*=.04, 95% CI =-.201, -.055) and the path via religious criteria (*b*=-.274, *SE*=.06, 95% CI = -.394, -.152). The total indirect effect of religiosity on evolution acceptance as mediated by religious and scientific criteria was -.392 (*SE*=.07).

Figure 4: Mediation Analyses

a.



b.



Notes: **p*<.05; ***p*<.01; ****p*<.0001. Mediation coefficients calculated using Hayes Process Model 4 (Hayes, 2013).

Discussion. Study 4 replicated the correspondence between constellation of criteria for belief accepted and acceptance of evolution found in Studies 1-3, indicating that it is indeed robust. More strikingly, acceptance of scientific or religious criteria outperformed religiosity and education as predictors for evolution acceptance. Moreover, the predictive power of criteria for belief was not restricted to evolution. Although evolution, as perhaps the most controversial, showed the strongest effect, criteria for belief also predicted acceptance of anthropogenic climate change and even the gaseous nature of the sun, a scientific claim not commonly considered controversial. This indicates that criteria for belief may play an important role in determining what people believe about science more broadly.

In addition, the mediation analyses are consistent with the hypothesis that the effects of religiosity and NoS on evolution acceptance are fully mediated by preferred criteria for belief. Yet we suspect that the true causal structure is far more complex. Preferred criteria are most likely only partial mediators of this effect. Other mediators, especially social pressure, are almost certainly also at play in the effects of religiosity. NoS and acceptance of scientific criteria are both likely to be higher in those who have received or sought out a more intensive education in science, which is likely to correspond with both greater social pressure and greater temperamental inclination to accept evolution. In addition, none of our variables are perfectly measured, and no observational study can definitively locate the causal arrows. Nonetheless, the mediation models suggest that criteria for belief may be an important and underappreciated mechanism by which social norms and education influence beliefs.

Study 4 found the strongest effects of the four studies. Perhaps this is because it included the most fine-grained measure of criteria for belief, asking participants to evaluate 20 candidate criteria on a 4-pt scale. It also included the finest-grained measure of evolution acceptance, allowing participants to respond on a 5-point scale to several different questions, including the

common descent of humans and apes as well as whether species came about via a "natural process." The ways in which people answer controversial questions can be complex, often combining aspects of both sides (Legare, Evans, Rosengren, & Harris, 2012). Acceptance of criteria for belief may also be complex. Further studies should continue to utilize relatively fine-grained measures, which capture a more nuanced range of beliefs and constellations of criteria.

General Discussion

Four studies explored which constellations of criteria for belief people think are legitimately justifying. These studies offered evidence that (a) such constellations of criteria differ substantially across individuals and groups, especially across differences of religion and understanding, and (b) these differences offer a novel explanation for the widespread rejection of evolution and other scientific theories. In all three surveys, the perceived legitimacy of different criteria for belief strongly predicted acceptance of evolution, supporting the patterns in the Study 1 interviews. The most critical criteria were the Bible, scientific evidence, and scientific consensus. Additionally, the perceived epistemic authority of parents and loved ones, the feeling of truth in the heart, and the conduciveness of a belief to ethical behavior predicted rejection of evolution. Accepting the criterion of epistemic coherence also predicted acceptance of evolution, perhaps because the scientifically-minded hold non-contradiction precious.

The predictive power of some criteria for belief persisted in regressions with known predictors religion, understanding of evolution, NoS, education, age, and gender, suggesting (though not conclusively) that the effect is not a mere artifact of affiliation or understanding. Moreover, several criteria for belief also predicted acceptance of climate change and the nature of the sun, showing the effect is not limited to evolution or even to controversial scientific claims.

This evidence demonstrates a robust empirical link between which criteria are taken as legitimate and the acceptance of scientifically justified claims. However, the current work is correlational and hence cannot conclusively demonstrate the hypothesized causal link between criteria for belief and beliefs. For this it would be necessary to manipulate the perceived legitimacy of criteria, which is likely to be extremely difficult. As noted in the introduction, these criteria are part of people's seldom-questioned cultural norms. Even if criteria were successfully manipulated, the effect of the criteria on beliefs might not occur immediately. It might also be necessary to induce reflection on justifications for long-held beliefs.

It is also probable that people apply criteria somewhat differently in different domains. That is, different criteria may be considered legitimate in evaluating candidate beliefs about people as opposed to physics. Additionally, it remains possible that these criteria are chosen *post hoc*, though not *ad hoc*, to support whatever beliefs one has. In any case, we have only just begun to investigate the relationship between beliefs and criteria for belief.

Nonetheless, these studies suggest that many people who hold counter-scientific beliefs may have acquired and confirmed them on the basis of reasons. The conflict arises because they take different reasons to be legitimate. That is, both those who accept evolution and those who reject it believe that one's views should be justified with respect to some criteria; a key part of the disagreement is over what those criteria should be. To communicate across these divides, we may need to discuss not only the presence of evidence and the consensus of particular groups, but also which criteria for belief we take as epistemically compelling, and why. There are reasons to think scientific criteria for belief are, at a minimum, better at tracking truth.

Conclusion

The practice of demanding justifications for beliefs is widespread. However, these studies indicate widespread disagreement on what counts as an adequate warrant. Many accept criteria which do not pass muster in the scientific community. Indeed, some put greater stock in non-scientific, personal criteria like the feelings of their heart or the word of their parents than in more objective criteria like scientific evidence. This emphasis on different criteria for belief makes different beliefs seem justified. Although the studies presented here focus on one controversial belief, the acceptance or rejection of evolution, our data suggests that accepted criteria may affect many of our beliefs. If scientists and educators are to persuade skeptics of the epistemic legitimacy of scientific theories, it is possible that no amount of augmenting the evidence or pointing to scientific consensus will be sufficient. It may be necessary to develop social norms which insist upon publicly accessible justifications for those beliefs which have more than merely personal consequences, like climate change and vaccines. It may also behoove educators and science communicators to explicitly address the particular power of scientific criteria for belief, if they wish science skeptics to heed them.

Bibliography:

Berinsky, A. J.; Huber, G. A.; & Lenz, G. S. (2012). Labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis, 20,* 351-368.

Bishop, B. A., & Anderson, C. W. (1990). Student conceptions of natural selection and its role in evolution. *Journal of research in science teaching, 27(5),* 415-427.

Brem, S.K., Ranney, M. & Schindel, J. (2003). Perceived consequences of evolution: College students perceive negative personal and social impact in evolutionary theory. *Science Education, 87,* 181-206.

Demastes, S. S., Settlage, J., & Good, R. (1995). Students' conceptions of natural selection and its role in evolution: Cases of replication and comparison. *Journal of Research in Science Teaching, 32*(5), 535-550.

Fleiss, J. L. (1971) "Measuring nominal scale agreement among many raters." *Psychological Bulletin, 76*(5), pp. 378–382

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19*(4), 25-42.

Funk, C. & Rainie, L. (January 29, 2015). Public and scientists' views on science and society. Pew Research Center. Retrieved 2/4/2016 from http://www.pewinternet.org/2015/01/29/public-and-scientists-views-on-science-and-society/

Gallup, Inc. (2014). Evolution, creationism, and intelligent design. Retrieved 22/7/2013 from http://www.gallup.com/poll/21814/evolution-creationism-intelligent-design.aspx.

Gervais, W. M. (2015). Override the controversy: Analytic thinking predicts endorsement of evolution. *Cognition, 142,* 312-321.

Gregory, T. R., & Ellis, C. A. (2009). Conceptions of evolution among science graduate students. *BioScience, 59*(9), 792-799.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* Guilford Press: New York.

Ingram, E. L., & Nelson, C. E. (2006). Relationship between achievement and students' acceptance of evolution or creation in an upper-level evolution course. *Journal of Research in Science Teaching, 43*(1), 7-24.

Johnson, R. L. & Peebles, E. E. (1987). The role of scientific understanding in college: Student acceptance of evolution. *The American Biology Teacher, 49, 2,* 93-98.

Kahan, D. M. (2015). Climate-Science Communication and the Measurement Problem. *Political Psychology, 36*(*1*), 1-43.

Kahan, D. M. (2016). 'Ordinary science intelligence': a science-comprehension measure for study of risk and science communication, with notes on evolution and climate change. *Journal of Risk Research*, 1-22.

Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition, 111(1)*, 138–143.

Kuhn, D. & Moore, W. (2015). Argumentation as core curriculum. *Learning: Research and Practice, 1(1),* 66-78.

Kuhn, D.; Zillmer, N.; Crowell, A.; & Zavala, J. (2013). Developing norms of argumentation: Metacognitive, epistemological, and social dimensions of developing argumentive competence. *Cognition & Instruction, 31*(4), 1-41.

Lawson, A. E., & Worsnop, W. A. (1992). Learning about evolution and rejecting a belief in special creation: Effects of reflective reasoning skill, prior knowledge, prior belief and religious commitment. *Journal of Research in Science Teaching, 29*(2), 143-166.

Legare, C.H.; Evans, E.M.; Rosengren, K.S.; & Harris, P.L. (2012). The coexistence of natural and supernatural explanations across cultures and development. *Child Development, 83 (3)*,779-793.

Lombrozo, T. A., Thanukos, A., & Weisberg, M. (2008). The importance of understanding the nature of science for the acceptance of evolution. *Evolution: Education and Outreach, 1,* 290-298.

McKay, R. T. & Dennett, D. C. (2009). The evolution of misbelief. *Behavioral and Brain Sciences, 32,* 493-561.

Metz, S. E.; Skolnick Weisberg, D.; & Weisberg, M. (2017). *Understanding of evolutionary theory is inadequate to motivate acceptance.* Manuscript in preparation.

Miller, J.D.; Scott, E.C.; Okamoto, S. (2006). Public acceptance of evolution. *Science, 313,* 765-766.

Motel, S. (September 23, 2014). Polls show most Americans believe in climate change, but give it low priority. Pew Fact Tank. http://www.pewresearch.org/fact-tank/2014/09/23/most-americans-believe-in-climate-change-but-give-it-low-priority/

Pennock, R. T. (2004). On teaching evolution and the nature of science. Chapter 1. In *Evolutionary Science and Society: Educating a New Generation*, Cracraft, J. & Bybee, R. W. (Eds.). Proceedings of the BSCS, AIBS Symposium, Chicago, IL.

Ryu, S. & Sandoval, W. A. (2012). Improvements to elementary children's epistemic understanding from sustained argumentation. *Science Education*, *93*(6), 488-526.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: new procedures and recommendations. *Psychological methods*, *7*(4), 422.

Shtulman, A. & Calabi, P. (2008). Learning, understanding, and acceptance: The case of evolution. *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, 235-240.

Simons, D. J. & Chabris, C. F. (2012). Common (mis)beliefs about memory: A replication and comparison of telephone and Mechanical Turk survey methods. *PLOS One, 7(12):* e51876.

Sinclair, A; Pendarvis, M. P.; & Baldwin, B. (1997). The relationship between college zoology students' beliefs about evolutionary theory and religion. *Journal of Research and Development in Education, 30 (2),* 118-125.

Stover, S. K., & Mabry, M. L. (2007). Influences of Teleological and Lamarckian Thinking on Student Understanding of Natural Selection.*Bioscene: Journal of College Biology Teaching*, *33*(1), 11-18.

Westfall, J. & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLoS ONE 11(3):* e0152719. doi: 10.1371/journal.pone.0152719

Williams, B. A. O. (2002). *Truth & truthfulness: An essay in genealogy*. Princeton University Press.

CHAPTER 2

IS EPISTEMOLOGY A STANCE OR A TOOLBOX? EPISTEMOLOGICAL REASONING IN ADOLESCENCE

Abstract:

The most widely accepted model of epistemological development (King & Kitchener, 1994; Kuhn, 1995; Perry, 1970) suggests that understanding of knowledge develops through a series of increasingly sophisticated *epistemic stages* in adolescence and early adulthood. The model posits that children begin with an *absolutist epistemic stage*, which allows only for objective aspects of knowledge; progress to a *multiplist epistemic stage,* wherein the objective is eclipsed by the subjective; and culminate in an *evaluativist epistemic stage*, wherein the subjective and objective aspects of knowledge are integrated. However, much of the research published within this tradition is empirically inconsistent with this model. I argue that epistemological development is better characterized as learning to draw upon and skillfully apply *epistemic technologies* from a toolbox, rather than a linear progression through stable epistemic stances. In a survey of 840 adolescents and interviews with 96 adolescents, predictions generated by the epistemic technologies model outperform those generated by the epistemic stance development model.

Introduction

How do we come to know? Epistemologists and philosophers of science have argued over this question for 2000 years. However, the vast majority of philosophers and scientists agree that any adequate theory of knowledge must take into account and integrate two aspects: (a) appeal to some form of shared reality, generally in the form of evidence; i.e., the role of the *objective*, as far as it goes; and (b) the underdetermination of theory by that evidence, making interpretation necessary; i.e., the role of the *subjective*. A solid grasp of both aspects is crucial for an adequate understanding of the nature of knowledge for effective reasoning.

The major paradigms of epistemological development all follow a similar pattern: they posit that children start out by appreciating only the objective aspects of knowledge (thinking only of the thing to be known). At some point, these theories postulate, people notice the importance of the subjective in interpretation (the role of the knower) and fall into an exaggerated

appreciation for the subjective aspects of knowledge (there is only the knower). Finally, some manage to bring the two aspects together for an integrated conception of knowledge (the knower gets epistemic access to the known by interpreting the evidence yielded by the thing to be known). In this paper I argue that this developmental model is oversimplified; the evidence demands a more fine-grained model of epistemological development. Noticing the subjective as well as objective aspects of knowledge occurs at a young age, and appreciation for subjective interpretation rarely dislodges all appreciation for the objective. The real challenge, I will argue, is integrating and applying both aspects of knowledge construction appropriately across domains and contexts.

Epistemological Stage Development Models. The widely accepted progression of epistemological development described above originated in Perry's (1970) study of epistemological and moral reasoning based on extensive interviews with Harvard undergraduates in the 1960s. In this sample, he commonly found both absolutism (which he called *dualism*) and multiplism (which he divided into *multiplicity* and *relativism*). Perry called the highest level of epistemological sophistication *commitment within relativism.* In his data, this emerged among only a minority of more advanced undergraduates. His progression was taken up and adapted by Baxter Magolda (1992, 2001, 2004) and further refined by King and Kitchener (1994, 2002) and Kuhn (1991, 1999, 2009; Kuhn, Cheney, & Weinstock, 2000).

The currently dominant family of paradigms, espoused by King, Kitchener, Kuhn, and many others, supposes that each individual makes sense of knowledge and justification through a dominant epistemic stance which changes as they become more sophisticated. The first epistemological stage begins in childhood as (a) simple objectivism about truth and reality, whereby knowledge corresponds straightforwardly to reality, limited only by mistakes or ignorance of facts; progresses through (b) a stage of multiplism or relativism in which everything

is a matter of subjective interpretation, making all beliefs equal; and finally, for some individuals, reaches (c) a final, sophisticated integration of the subjective and objective aspects of knowledge construction. This last stage is known variously as *evaluativism* (Kuhn et al. 2000), *post-skeptical rationalism* (Boyes & Chandler, 1992), *metasubjectivist rationality* (Moshman, 2015), or *integration* (Mansfield & Clinchy, 2002), to name just a few. Different theorists have emphasized different aspects of this final stage, but they all emphasize the appropriate integration of the subjective and objective aspects of knowledge construction. Different research paradigms within this tradition sometimes split the stages into sub-stages or mark the divisions in slightly different places, but this basic three-step progression has won a general consensus among many scholars (Baxter Magolda, 1992, 2001, 2004; King & Kitchener, 1994, 2002; Krettenauer, 2004, 2005; Kuhn, 1991, 1999, 2009; Kuhn, Cheney, & Weinstock, 2000; Kuhn & Park, 2005; Mason, Boldrin, & Zurlo, 2006; Mason & Scirica, 2006; Olafson, Schraw, & Vander Veldt, 2010; Weinstock, Neuman, & Glassner, 2006).

King and Kitchener (1994, 2002) and Kuhn (1991, 2000) have argued that these three stances – absolutism, multiplism, and evaluativism – form a developmental trajectory in the sense that any given individual reasons either within one stage or within two adjacent stages, and the stages seem to form a more or less invariant developmental sequence. Some never move beyond the absolutist stance, while others become stuck in the multiplist stance.

Kuhn suggests that a shift to the multiplist stance is sparked by a recognition that even experts can disagree, which draws reasoners into skepticism of epistemic authority of any kind. Multiplists reason that if even experts can disagree, there are no grounds to take them as epistemic authorities. An evaluativist stance requires the recognition that some methods of obtaining knowledge are objectively better than others, even if none are infallible.

The family of epistemological stage development (ESD) models has the virtues of simplicity and fit with much of the evidence. Many studies have found that absolutism decreases and multiplism and evaluativism increase across the span of adolescence and college (King & Kitchener, 2002; King, Kitchener, & Wood, 1994; Kuhn, 1991; Mansfield & Clinchy, 2000). As further corroboration, a number of studies have found links between dominant epistemic stance and reasoning quality in a variety of contexts (Kuhn, 1991; Mason & Boscolo, 2004; Mason & Scirica, 2006; Weinstock, Neuman, & Glassner, 2006). For instance, among prospective jurors, low epistemic stance predicted premature judgment, while an evaluativist epistemic stance outpredicted education, age, and confidence on a measure of reasoning quality (Weinstock, 2005; Weinstock & Cronin, 2003). Among adolescents, a more sophisticated epistemic stance has also predicted the development of a considered and coherent identity (Krettenauer, 2005).

Following Kuhn, such a link between epistemic stance and reasoning quality is exactly what we should expect. Epistemic stance constitutes an individual's conception of the nature of knowledge and how we know. As such, epistemic stance should shape intellectual values and practices. Both absolutist and multiplist stances undermine the motives for careful, unbiased epistemic inquiry. A pure absolutist believes that knowledge can be acquired straightforwardly from reading it more or less directly off reality or else by consulting the appropriate authorities. The absolutist subject hence supposes herself already in possession of the truth or at least able to acquire straightforward access to it, rendering her beliefs dogmatic and oversimplifying approaches to inquiry. Meanwhile, multiplists believe knowledge is entirely relative and any belief is as good as another, which implies that an objective truth beyond our perceptions is either wholly inaccessible or else nonexistent. This belief denies the possibility of rational comparison of candidate theories and thus discourages the attempt.

Both of these epistemic stances encourage pessimism about the utility of collecting more evidence or considering more alternatives. Absolutism calls for rigidity. Multiplism calls for passivity. Only those who think that knowledge is uncertain yet not wholly out of our reach have reason to persist in examining the knowledge we think we have and seeking that which we lack. Indeed, Deanna Kuhn (2001) has argued, "Only [the] most advanced levels [of epistemological understanding] support a disposition to engage in the intellectual effort that reasoned argument entails" (p. 1). If this argument is right, then epistemological sophistication is crucial for the critical, actively open-minded thinking on which a well-functioning democracy depends.

Despite the virtues and wide agreement on the ESD model, it struggles to explain two well-established phenomena: (a) the domain specificity of epistemic stance and (b) stark inconsistencies in the research literature for when an appreciation for interpretation arises and when integration of the objective and subjective occurs.

Domain Specificity of Epistemic Stance. Researchers refer to the subject area of an epistemological question as its domain, distinguishing, for example, the domain of judgments about aesthetics from that of judgments about social science. Most of the proponents of the ESD model acknowledge that progression through the three stages can happen more quickly in some domains than others, leading to some differences in epistemic stance across domains (Kuhn, 2009; Kuhn, Cheney, & Weinstock, 2000; King & Kitchener, 2002). Kuhn herself uses measures of epistemological understanding in different domains, and, like others, has frequently found more multiplism in domains of taste and aesthetics and more absolutism or evaluativism in domains of natural and social science (Kuhn, 1999, 2009; Kuhn, Cheney, & Weinstock, 2000; Mason, Boldrin, & Zurlo, 2006). Nevertheless, these scholars still tend to take the dominant epistemic stance as primary, and most domain differences as temporary fluctuations around a central epistemic stance as individuals move through the progression in each domain (with

possible exceptions for the domains of taste and aesthetics, which frequently stop at the multiplist level) (Kuhn, Cheney, & Weinstock, 2000).

However, other researchers think domain specificity is itself the dominant pattern. Muis, Bendixen, and Haerle (2006) reviewed 19 studies of this question, and found that 17 of them offered stronger evidence for domain specificity than domain generality. Appreciation for the role of interpretation and presence of uncertainty are commoner in the social sciences and humanities than in the hard sciences, while appreciation for the role of evidence is more common in the hard sciences. However, eight of the studies offered evidence for some degree of domain generality. Muis et al. conclude that epistemological beliefs can differ substantially across domains, but there remains some domain generality. Hofer (2006a, 2006b) agrees with Muis et al., further distinguishing between *enacted epistemological beliefs*, which shape reasoning in particular contexts and domains, and *espoused epistemological beliefs*, which shape responses to direct questions about the nature of knowledge, be they domain-general or domain-specific.

Hofer (2006a, b) also suggests that *context*, i.e. exactly how epistemological reasoning is elicited, affects both enacted and espoused epistemological beliefs. Hofer and others argue that a good theory of epistemological development must account for both patterns of domain specificity and domain generality, while leaving room for other contextual factors. Evidence from studies of meta-ethical reasoning supports this; for example, participants judge moral judgments as more or less objective depending on the similarity of the moral agents to the themselves (Sousa, Piazza, & Goodwin, 2016).

A version of the ESD model might have the resources to do so. Moshman (2015) seeks to account for domain specificity by suggesting that epistemological development occurs separately in each domain, with enough transfer insights to yield a moderate degree of domain-generality.

He also submits the possibility that evaluativism may not be normative for all domains; for instance, in sufficiently well-defined matters of arithmetic or logic, an absolutist stance may be not merely legitimate but optimal. If we take Moshman's suggestions seriously, we should adopt a modified version of the ESD model with much more domain specificity. However, domain specificity is not the only problem with the current version of the ESD model.

Age-Related Differences in Epistemic Stance. There is considerable disagreement in the literature concerning when an appreciation for the role of interpretation arises, as well as when integration of objective and subjective aspects of knowledge-construction appears. The dominant paradigm suggests that both progressions occur primarily during adolescence (Kuhn, 1991; Mansfield & Clinchy, 2000, Moshman, 2015) or college (Baxter Magolda, 2004; King & Kitchener, 2002). Some studies indicate that integration of the subjective and objective is rare among undergraduates (Baxter Magolda, 2004; Perry, 1970). However, others indicate that interpretation and even integration appear as early as age seven or eight (Carpendale & Chandler, 1996).

Perry (1970) and Baxter Magolda (2004) conducted extensive interviews with undergraduates, which revealed only occasional integration of the objective and subjective aspects of knowing. In Perry's sample of Harvard undergraduates, he found both absolutism (which he called "dualism") and multiplism more common than a full integration of the two.

In Baxter Magolda's (2004) longitudinal interview study of 101 undergraduates, she counted only one third of entering freshmen as showing signs of anything more complex than absolutist knowing. These students still described knowledge as absolute and certain in some domains, but not all. Baxter Magolda called this stage *transitional knowing.* These students acknowledged that there are some domains in which there isn't a clear right answer or a clear and

reliable way to find the right answer. Yet even among these transitional students, many described this property of uncertainty in order to complain about it. One said, "I feel uneasy in classes like that." Another insisted, "There have to be more guidelines or structure. You have to have something more firm here; there's never a straight answer."

It seems that an appreciation for subjectivity in taste judgments starts early and sticks. Children as young as three commonly recognize that conflicting tastes, e.g. of food, may be equally valid (Flavell, Flavell, Green, & Moses, 1990). Children as young as 18 months can accurately guess which food a confederate will prefer based on that confederate's past behavior, even when it contrasts with their own preferences (Repacholi & Gopnik, 1997). Reasoning about the personal taste domain, however, may be unrepresentative. Multiplism is much the commonest epistemic stance in this domain at all ages tested, with absolutism quite rare (Kuhn, Cheney, & Weinstock, 2000; Mason, Boldrin, & Zurlo, 2006).

Carpendale and Chandler (1996) agree with the ESD model that very young children do not appreciate the role of interpretation in knowledge construction; they lack the insight that "the 'mind' influences how the 'world' is experienced" (p. 1688). However, their studies place the arrival of this insight about five years earlier than the classic paradigm. They developed a set of tasks for young children using puppets which disagreed with each other about ambiguous visual or verbal stimuli. For instance, presented with the iconic duckrabbit figure (a drawing that can be seen as either a duck or a rabbit), one puppet said it was a duck, the other said it was a rabbit. Children ranging from 5 to 8 years were asked why the puppets disagreed, and what a third person might think. The younger children said only one puppet could be right, consistent with the DEP. However, the majority of children over seven not only said both puppets were right, but also successfully attributed their disagreement to the ambiguity of the stimulus, thereby acknowledging a role for subjective interpretation.

This very idea that both of two disagreeing agents could be right is precisely the same concept used to indicate a non-absolutist epistemic stance in Kuhn, Cheney, and Weinstock's (2000) widely used measure of epistemic stance (Cho, Lee, & Jonassen, 2011; Mason, Boldrin, & Zurlo, 2006; Mason & Boscolo, 2004; Mason & Scirica, 2006; Weinstock, Neuman, & Glassner, 2006). Still more strikingly, 77% of children over age seven also said that a third puppet that called the drawing an elephant was wrong. This suggests that these young children were not multiplists (for whom all three opinions should be equally good), but *evaluativists.* At a minimum, the children understood that reasonable interpretation was constrained partly but not completely by the evidence, indicating some degree of awareness of both subjective and objective aspects of knowledge.

Similarly, Wainryb et al. (2004) found that some five-year-olds and most seven and nine-year-olds took some beliefs as more relative or subjective than other beliefs. Moreover, nearly all justified judgments that some beliefs were not relative by appeal to truth. This indicates that not only are young children able to apply subjective epistemological reasoning, but further, they can appropriately differentiate cases on which it is possible to be more and less objective.

Kuhn, Cheney, & Weinstock's (2000) interviews with 21 second and third graders also indicated the early onset of multiplist reasoning, although they did not interpret their data thus. According to their schema, eight children showed multiplist reasoning on some questions and absolutist reasoning on others, indicating some grasp of both objective and subjective aspects of knowledge though no integration. Multiple children (the numbers are unspecified) were coded as consistent multiplists and several as edging into evaluativism. Contrary to Kuhn's later claim that "Progression beyond absolutism is unlikely during childhood," (Kuhn, 2009, p. 113) she and her colleagues coded less than half of the second and third graders in this sample as absolutists based on their interviews.

Can the ESD model be modified to explain such drastic developmental inconsistencies? Proponents have made several attempts to accommodate it. Kuhn (2009) has suggested that although the recognition that different minds may interpret reality differently can arise in early elementary school, the belief that only one such interpretation could be correct persists until adolescence. However, this does not account for Chandler and Carpendale's (1996) cases, in which many children say both interpretations are right.

Moshman (2015) has suggested that epistemological development takes a spiral shape, with capacities for greater epistemological sophistication first developing in concrete and familiar cases through childhood, then the same pattern re-occurring as a more general and abstracted epistemic stance develops through adolescence and adulthood. This solution has a certain elegance and could explain the children's success with duckrabbits. However, if epistemological reasoning becomes increasingly general and abstract, then domain specificity should decrease through adolescence. Yet reviews indicate that domain specificity persists through adolescence and adulthood (Muis, Bendixen, & Haerle, 2006; Moshman, 2015).

The wide disparities in the results of interviews with young children and adults may be largely due to wide variation in the questions asked and the coding schemata. The undergraduates presumably would have also appreciated the ambiguous nature of the duckrabbit, and in that context shown at least as much awareness of subjective and objective aspects of knowledge as the children. It appears that the form and content of questions can elicit quite different forms of epistemological reasoning.

The evidence that even fairly young children have the capacity to appreciate some role for interpretation in knowledge construction is strong. However, there is still considerable room for development. It is still possible that a single, sufficiently clear rubric of epistemological

development through the stances would show the gradual development through adolescence and early adulthood described by the ESD model.

Fortunately, a number of studies have been conducted by the proponents of the ESD model using the same written measure across age groups, facilitating comparison. The most widely used quantitative measure of epistemic stance is the 15-item measure developed by Kuhn, Cheney, and Weinstock (2000). It includes three items in each of five domains: natural science explanation, social science explanation, moral claims, aesthetic claims, and taste. Because responses across items and domains are often inconsistent, individuals are assigned the epistemic stance indicated by their modal response.

The study which introduced this measure (Kuhn, Cheney, & Weinstock, 2000) included 20-25 participants from each of six populations: fifth graders, eighth graders, twelfth graders, undergraduates at an elite university, community college students, and MBA students at a major business school. Although there were some developmental trends consistent with the ESD model, they were slight and inconsistent even over this decade of development. Fifth graders were somewhat more likely to be absolutists (30%) about natural science than the older groups (10%-22%), but eighth graders were already relatively unlikely to be absolutists about anything (12% in the domain of natural science explanations, 16% in the domain of value, zero for everything else). The frequency of evaluativism across domains was only significantly different between fifth graders and undergraduates (p=.041). But perhaps these weak effects are an artifact of the small sample size at each age.

Focusing in on the empirical truth domains (natural science and social science explanations), Weinstock, Neuman, and Glassner (2006) gave the same measure to 189 7th, 9th, and 11th graders in Israel. In this population, the absolutist epistemic stance was actually more

common in older students – 8% of 7th graders vs. 22% of 11th graders – while incidence of evaluativism increased insignificantly from 28% to 30% across the four-year age gap.

Mason, Boldrin, and Zurlo (2006) gave the same measure to 881 Italian students in 5th grade, 8th grade, 11th grade, and 13th grade. As in Kuhn, Cheney, and Weinstock's (2000) sample, there was a significant decrease in absolutism from fifth grade to older grades (p=.01). However, this does not indicate a clear cross-domain developmental trend: this difference arises almost entirely from the fact that fifth graders are more likely to be absolutists about value than older students. Both eighth graders and thirteenth graders at the technical school were exactly as likely as fifth graders to be absolutists about natural science explanation (22%). In addition, rates of evaluativism averaged across domains were virtually identical for fifth graders and 11th and 13th graders at the higher-level scientific high school (respectively, 22.2%, 22.6%, and 22.2%).

Recall that this eight-year period – age 11 to age 19 – is exactly the period in which the ESD model predicts the greatest amount of epistemological development. Even if epistemological development were linked wholly to education and not to development, one would expect to see more progress across this key period. It would be an exaggeration to say there is zero evidence of a developmental trend away from absolutism and towards evaluativism. However, the weakness of these developmental trends should cast serious doubt on the currently dominant picture of a developmental progression from absolutism to multiplism to evaluativism.

Furthermore, there is some evidence that education may play a role independent of development. In the domain of moral judgments, years in college predicts lower rates of moral absolutism, both among enrolled undergraduates (Nichols, 2004) and in contrast to other, less educated adults as well as children (Beebe & Sackris, *in press*).

In short, the dominant paradigm of ESD has two major shortcomings. It does not adequately account for either the large degree of domain specificity nor the low degree of age-related epistemological development supported by the evidence. On the other hand, it does offer an account of three forms of epistemological reasoning of varying sophistication which seem to map approximately onto a large number of qualitative interviews. I suggest that another theoretical account of epistemological development may be able to build on the insights of the ESD model to account for all of these findings.

The Epistemic Technologies Model. I suggest that it would be more fruitful to approach epistemological thinking as a process of drawing upon one's repertoire of related *epistemic technologies*, conceptual tools for building and evaluating representations that aim at knowledge. Each thinker has a finite set of epistemic technologies in their repertoire, some more accessible than others, and often tagged as appropriate for particular contexts. These technologies can be (a) *epistemic forms*, ways of structuring representations to make them easier to examine, e.g. "lists" or "bar graphs"; (b) *epistemic strategies*, actions to build or evaluate candidate beliefs, e.g. "consult my intuition," "consult a textbook," or "collect relevant evidence"; or (c) *representational modes*, like "know," "doubt," "be 90% sure," "perceive," "conceive as," or "opine."

This ontology is meant to be a starting point, not necessarily an exhaustive list, and is not created *de novo.* The idea of epistemic technologies is very similar to Hammer and Elby's (2002, 2003, 2010; Louca, Elby, Hammer, & Kagey, 2004) theory of *epistemological resources*, which, as here, focuses on fine-grain sized elements of epistemological understanding which are drawn upon differently in different contexts as well as different domains. The notions of *epistemic forms* and *epistemic strategies* are drawn from Collins & Ferguson (1993), who call strategies *games*.

Epistemic technologies are generally first adopted in their simplest versions, and not immediately applied with perfect skill. Thus it is possible for fairly young children to grasp a great many epistemic technologies, but still have plenty of room for improvement. The sophistication of a given epistemic technology can develop along three dimensions: (a) developing the simple version into more complex, sophisticated versions (e.g. from *lists* to *outlines*; from *try it and see what happens* to *randomized controlled trial*; or from *mistaken because something wasn't seen* to *mistaken because something wasn't noticed*); (b) learning to recognize more contexts in which it is and is not applicable (e.g. realizing that not only peers but also teachers can be mistaken); and (c) learning to apply it more effectively (e.g. recognizing the need in an RCT to double-check that a control group is sufficiently similar to an experimental group).

The classic ESD model has a Hegelian form: thesis (knowledge is objective), antithesis (knowledge is subjective), synthesis (knowledge is constructed on both objective and subjective bases). Contrary to the ESD model, the epistemic technologies model suggests that developing knowers may not move through these three as separate temporal stages. Instead, a full understanding of the objectivity and subjectivity of knowledge is composed of a number of more basic epistemic technologies, most of which are accessible even to fairly young children. The epistemic technologies which make up an appreciation of the objectivity of knowledge include epistemic strategies like *appeal to evidence* (where evidence is what has been observed, without regard for the role of the observer). Epistemic technologies which make up an appreciation of the subjective aspects of knowledge construction include representational modes like *this is one interpretation* and *seems plausible to me but maybe not to others*, and beliefs like *multiple interpretations can be reasonable*.

If the epistemic technologies model is more accurate, then we should expect to see objective and subjective epistemic technologies coexisting even among fairly young children; we should see even individuals who talk like paradigmatic multiplists using objective epistemic technologies in some contexts; and we should see some individuals going from the overemphasis on the objective classic in childhood to a proper integration of the objective with the subjective without ever passing through a phase of excessive emphasis on the subjective.

These predictions fit fairly well with the available evidence. For instance, both objective appeals to evidence and appreciation for multiple subjective interpretations are common among elementary school children (Chandler & Carpendale, 1996; K. Metz, 2004; Flavell, Flavell, Green, & Moses, 1990). With practice, children and adults also become increasingly good at identifying contexts in which each epistemic technology is applicable. A big part of that increased sophistication is learning how to integrate the subjective and objective aspects of knowledge construction.

Even among those who possess highly sophisticated versions of particular epistemic technologies, we should expect varied sophistication in the application of those technologies. An abundant literature shows that transfer of ideas and techniques across contexts and domains is extremely difficult (Brown & Kane, 1988; Detterman, 1993; Gick & Holyoak, 1983; Holyoak, 2012; Novick, 1990). Even college students often fail to transfer solutions across very similar problems without explicit cues (Reed, 1987). However, the greater the number of contexts a subject successfully applies a given approach, the easier it becomes to apply that approach in a new context (Gick & Holyoak, 1983; Holyoak, 2012).

On account of these known challenges of transfer, epistemic technologies theory predicts substantial domain specificity in epistemic stance, as is observed. However, the possibility of

transfer predicts that there will be some degree of domain generality, such that once someone know how to apply an epistemic technology in one domain, it becomes easier to extend that same technology to other domains. Moreover, the more domains in which an idea is applied, the easier it should become to apply it in further domains. For instance, someone who has detected measurement error in a yardstick, an odometer, and a thermometer is more likely to consider the possibility of measurement error on a scale than someone who has only noted it with yardsticks. Successful transfer to one novel domain aids abstraction of the underlying principle, making transfer to further domains increasingly easy (Gick & Holyoak, 1983; Holyoak, 2012).

If epistemological reasoning is based on piecewise epistemic technologies used only some of the time, then contextual factors should heavily influence epistemological reasoning. The epistemic technologies model thus takes the effects of context as crucial, while the ESD model takes them as a distraction from true competency. According to the epistemic technologies model, the greater the competence, the fewer cues will be needed to recognize that both subjective and objective factors are relevant. This makes sense of the striking findings discussed above: namely, that contextual cues can make an appreciation for the subjective aspects of knowledge turn up as young as seven and shirk as late as twenty.

In this way, the epistemic technologies model can also explain the irregular evidence for developmental progress. On an epistemic technologies model, observed differences in epistemological sophistication largely arise from interpretation of the question, domain, and education. Differences in reasoning may also arise from the extent to which existing epistemic technologies have been elaborated and integrated by the subject.

I concede that some individuals tend to draw upon objectivist epistemic technologies, while other individuals tend to draw upon subjectivist epistemic technologies. However, I

contend that *there is no phase in school-age development of either complete objectivism or of complete subjectivism.* Most five-year-olds already appreciate the role of subjectivity in matters of taste (Flavell, Flavell, Green, & Moses, 1990), though perhaps not matters of natural science. Meanwhile, radical subjectivism is not a position which can be sustained in ordinary life, where appeals to some kind of shared reality are necessary for mundanities from showing up on time to communicating a shrimp allergy. The evidence suggests that, contrary to ESD, subjective and objective epistemic technologies coexist in thinkers from a young age.

So much of the existing research on epistemological development has been interpreted and presented with the ESD in mind that it is difficult to tell if children and adolescents really can skip the multiplist phase. For longitudinal studies which observe absolutism at one time point and evaluativism at the next, it is easy to suppose that the multiplist phase occurred in between interviews. However, if a large number of apparent multiplists can be observed applying objective epistemic technologies, that is good reason to think that the true multiplist is a fictional species. Indeed, I argue that this is exactly what we see; those who insist that all beliefs are equally good are no more likely than others to think that an allergy to shrimp can be cured by believing the shrimp a vegetable or that one can will oneself on time by believing without evidence that the classroom clock is ten minutes fast.

The challenge facing both children and adults is to apply these technologies in a coherent and effective way to make sense of the world. As is acknowledged by all researchers in the psychology of epistemology, this task is not one which most adults have mastered. Indeed, professional scientists at prestigious universities have made mistakes in the coordination of theory with evidence, underestimating the role played by theory and overinterpreting ambiguous and even disconfirming data by focusing on the pieces that do fit (Simmons, Nelson, & Simonsohn, 2011; Simonsohn, Nelson, & Simmons, 2014). I offer evidence here that many of the epistemic

technologies utilized by professional scientists are also in the repertoire of young adolescents, albeit drawn upon with less skill and in a less integrated fashion. Insofar as this is the case, a piecemeal approach to epistemological development may be more fruitful for educators than an attempt to push children from absolutism to multiplism at the expense of appreciation for evidence and objectivity.

Overview of Studies. In the present study, I offer two lines of evidence that the epistemic technologies model accounts for better than the ESD model. First, I add to the already extensive evidence of domain and topic specificity in epistemological reasoning, based on the widely used measure developed by Kuhn, Cheney, & Weinstock (2000). Second, I test a pair of competing predictions from the epistemic technologies account vs. the ESD model: the former predicts that subjectivist and objectivist epistemic technologies will often coexist in young adolescents, while the ESD model suggests such coexistence would arise mainly in evaluativists and hence will be relatively rare at this age. The phenomenology of adolescent epistemological reasoning is explored through the lenses of the epistemic technologies model and the ESD model.

Study 1: Survey of Epistemic Stance

In the first study, I sought to investigate inconsistencies in expressed epistemic stance. According to the ESD model, inconsistencies in expressed epistemic stance are largely noise, with the exception of some domain specificity. According to the epistemic technologies model, however, we should expect to see considerable inconsistency in expressed epistemic stance not only across domains, but also within domains and even between different forms of the same question. To investigate the extent of such hypothesized inconsistencies, the most widely used written measure used by proponents of the ESD model, was administered to a large, diverse sample of adolescents (Kuhn, Cheney, & Weinstock, 2000).

Method.

Population. The epistemic stance measure was given to virtually all ninth graders at three high schools across the United States, including 29 students from a private school, 140 students from a charter school, and 672 students from a public high school, for a total of *N*=841 students. This sample was 51.1% female, with an average age of 14.9 years (*SD*=.44). It was ethnically diverse, composed of 39.4% African-American, 27.5% Caucasian, 17.8% Hispanic, 12.7% Asian-American, 1.1% multiracial, and 0.4% Native American students. English language learners made up 15.6% of the sample, and 63.6% qualified for free or reduced lunch.

Procedure. Demographic questions and an epistemic stance measure were administered as part of a larger survey to students in class on school computers using the Qualtrics survey platform. An opt-out consent procedure allowed almost all students to participate. Teachers and/or research assistants were on hand to answer any questions or fix technology issues. The whole survey, which included many other items, mostly rating-scales, took an average of 45 minutes to complete. Participants had completed similar surveys up to three times before across the previous 18 months, but none had seen the epistemic stance measure before.

Materials. Participants completed an eleven-item version of the measure of epistemic stance developed by Kuhn, Cheney, & Weinstock (2000). The original version included three items in each of five domains: natural science, social science, moral claims, aesthetic claims, and personal taste. Because the last two domains tend to have different properties and the current interest is primarily in epistemological reasoning about scientifically investigable topics, only the two scientific domains were included here, plus two items from the moral domain for comparison.

Two characters, Robin and Chris, are introduced who disagree in each item about which of two accounts is right (See Table 5). Each item then consists of two parts. First, subjects are

asked if "both views could have some rightness," or "Only one of their views could be right." After they have answered this question, those who said both could be right are asked to select either "One could be more right," or "One could not be more right than the other." Those who answer the first question by saying only one could be right are scored as Absolutists for that item; those who say neither could be more right are scored as Multiplists; and those who say one could be more right than the other are scored as Evaluativists.

Table 5

Epistemic Stance Multiple Choice Items

	Target of Disagreement
Natural 1	R believes one book's explanation of what atoms are made of. C believes a different book's explanation of what atoms are made up of.
Natural 2	R believes one book's explanation of how the brain works. C believes a different book's explanation of how the brain works.
Natural 3	R believes one mathematician's proof of the math formula is right. C believes a different mathematician's proof of the math formula is right.
Natural 4	R believes one explanation of how dinosaurs went extinct. C believes a different explanation of how dinosaurs went extinct.
Natural 5	R believes one explanation for why earthquakes happen when and where they do. C believes a different explanation for why earthquakes happen when and where they do.
Natural 6	R believes the universe will last forever. C believes the universe will not last forever.
Social 1	R has one view of why criminals keep going back to crime. C has a different view of why criminals keep going back to crime.
Social 2	R thinks one book's explanation of why the Crimean wars began is right. C believes a different book's explanation of why the Crimean wars began is right.
Social 3	R agrees with one book's explanation of how children learn language. C agrees with a different book's explanation of how children learn language.
Moral 1	R thinks the government should limit the number of children families are allowed to have, to keep the population from getting too big. C thinks families should have as many children as they choose.
Moral 2	R thinks lying is wrong. C thinks lying is permissible in certain situations.

In past applications of this measure, those who answer the first question as Absolutists are not given a chance to answer the second question. Here, all participants were asked to answer both questions, regardless of their answer to the first question, and a third option was added to the second question, with a tag to acknowledge the redundancy: "I said only one could be right." This format permitted the questions to be scored in the same way as previous applications, by scoring as absolutists those who answered the first question, "Only one of their views could be right" and scoring the second question only for those who first said, "Both views could have some rightness." It also allowed the detection of individuals who changed their mind when faced with a more nuanced set of options. If elements of context like available options influence epistemological reasoning, as Hammer and Elby (2002, 2006) have argued, some participants should change their mind and respond inconsistently across the two items.

Given the study's focus on the development of scientific understanding, I added three items on natural science to the survey, one (about earthquakes) paralleling a question asked in the interview section. The original wording of the natural science domain questions opposed the explanations given in two different books. It was thus merely implied, and not stated, that the explanations were different, although the modal form of the questions ("*Could* both be right" and "*Could* one have more rightness") should be sufficient to encourage participants to consider explanations that *were* different. The novel items referred to "different explanations" instead of different books' explanations in order to make more explicit the intended contrast between explanations. Further, because different explanations are more likely to turn out to be compatible than directly opposing claims, the third item added described directly opposed but still debatable natural science claims: "Robin believes the universe will last forever. Chris believes the universe will not last forever." This was intended as an explorative item, to see if epistemic stance about

claims matched up with epistemic stance about *explanations.* The six natural science items, three social science items, and two moral items added up to a total of eleven items.

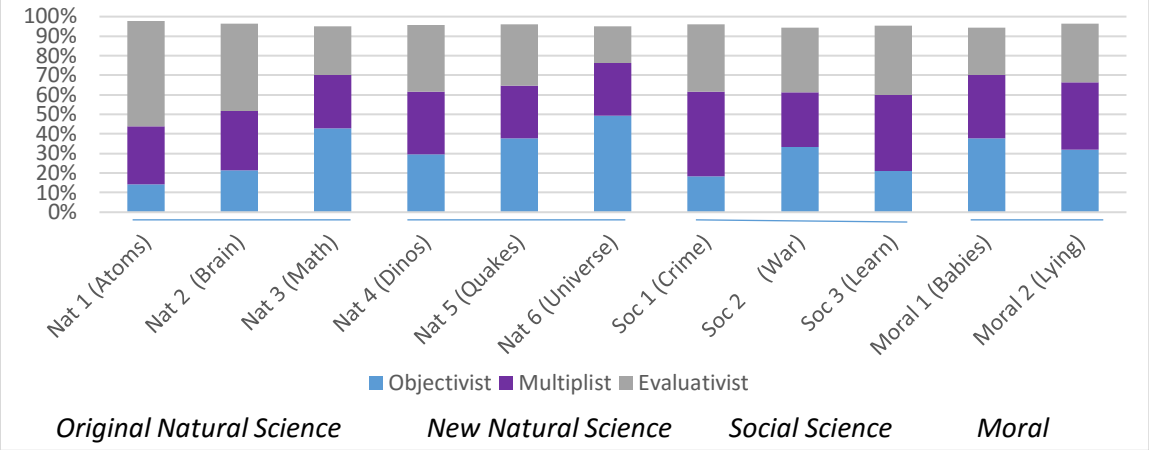
Results.

Item level frequencies. The items with the highest proportion of evaluativist responses were both in the domain of Natural Science (Nat). Nat 1, about competing explanations of what an atom is made of, elicited the most evaluativism at 53.9%, with 29.6% multiplist responses and only 14.2% objectivist responses. Nat 2, competing explanations for how the brain works, was the next most likely to elicit an evaluativist response (44.5%), with 30.2% multiplist and 21.5% objectivist responses. However, the two items which elicited the most objectivist responses were also both in the Natural Science domain: Nat 3 (competing mathematical proofs: 42.9% objectivist responses) and Nat 6 (will the universe last forever: 49.4% objectivist responses). This is not especially surprising, since math is often taken as the epitome of objectivity and is not a proper natural science at all[2], while the immortality of the universe is the only item phrased in terms of directly opposing claims rather than competing explanations. Even so, 28.6% gave the multiplist response to the Universe item, saying that it is equally right that the universe will last forever and that it will end. Evaluativism seems to be more commonly acknowledged for competing explanations than competing claims, though the items are not matched closely enough to show this conclusively.

By and large, items from different domains did not elicit substantially different patterns of responses, although particular items did somewhat (see Figure 5.) This somewhat undermines the idea of domain specificity in epistemological reasoning, although not of topic specificity.

[2] Nat3 is categorized with the Natural Science items following Kuhn, Cheney, & Weinstock (2000).

Figure 5. Frequencies of Epistemic Stance by Item



Note: Frequencies do not quite sum to 100% because of missing data.

Domain-level frequencies. Kuhn, Cheney, and Weinstock (2000) and most of those who used the measure thereafter coded epistemic stance in each domain by taking the most frequent response of a given subject in that domain. Since there are three items per domain, this means an epistemic stance indicated more than once is the mode. However, it is also possible for subjects to indicate a different epistemic stance on each of the three items.

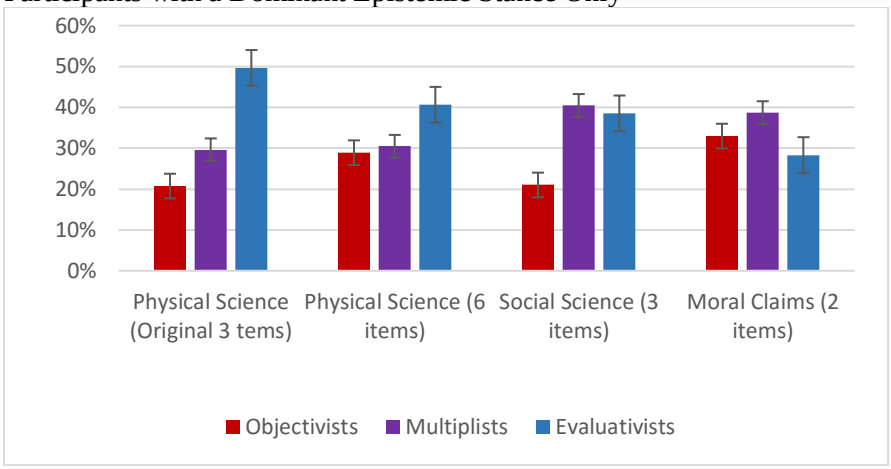
Analyzing the data in this way, 86.4% of respondents did exhibit a dominant epistemic stance within the natural science domain (based on the three original items) and 88.7% within the social science domain. However, if the three additional natural science items are added to the original three, only 54% of respondents had a majority answer across those six items.

The proportions of each epistemic stance within each domain are shown in Figure 6. The original three natural science items elicited 17% absolutists, 26% multiplists, 43% evaluativists, and 14% with no dominant epistemic stance (a different answer for each item). On the six natural science items, 16% gave majority absolutist answers, 16% multiplist answers, 22% evaluativist

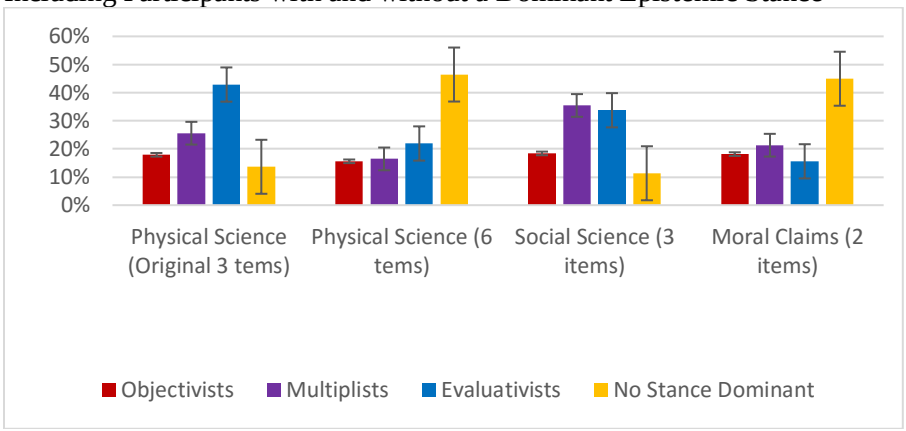
answers, and 46% did not give any one answer on more than half the items. In the social science domain, 18% counted as absolutists, 35% as multiplists, and 34% as evaluativists, while in 11% no one epistemic stance predominated. The higher rates of evaluativism in natural science and of multiplism in social science are consistent with past research (Kuhn, Cheney, & Weinstock, 2000; Moshman, 2015).

Figure 6: Frequencies of Epistemic Stance by Domain

a. Participants with a Dominant Epistemic Stance Only



b. Including Participants with and without a Dominant Epistemic Stance



Note: Error bars show standard error.

Because many respondents were not sufficiently consistent within domains to be assigned one epistemic stance, I show results both including and excluding these individuals. Figure 6a shows proportions of those in each domain who could be assigned an epistemic stance, not

counting those who could not. Figure 6b shows proportions of the whole sample including those who did not indicate a single dominant epistemic stance.

Note that the plurality of those with any dominant stance in natural science are evaluativists (more evaluativists than multiplists: t=6.14 (708), p<.001; more evaluativists than absolutists: t=9.77(708), p<.001). Multiplism and evaluativism are about on par for social science (t=.59 (718), *ns*), both commoner than absolutism (t=6.87 (718), p<.001). Moreover, the reduced number of participants with any dominant epistemic stance from the three-item to the six-item measure in natural science primarily reduces the frequency of evaluativists (reduced by 21%), reducing multiplists by only 9% and absolutists by 2%. This suggests that the incidence of evaluativism could be underestimated by this measure, since predominant evaluativists are more likely to give different answers on different items. Indeed, such differentiated responses to different items arguably indicate a form of sophistication, since on some interpretations of the items, non-evaluativist responses could be normative.

Consistency of epistemic stance. There was considerable inconsistency even within domains of expressed epistemic stance. As can be seen in Figure 6b, a plurality of participants did not show any dominant epistemic stance in either the six-item natural science domain set or the two-item moral domain set. Although some of this is certainly due to noise, these inconsistencies—which arose despite the close similarity of the items' structure—are enough to support the thesis that some topics are more likely than others to cue particular epistemic technologies.

A clearer measure of within-domain consistency is how often respondents consistently endorsed one epistemic stance within a domain. In the natural science domain, 30% were consistent across the first three items, and only 11% were fully consistent across all six items. In

the social science domain, 40% were consistent across all three items. These rates indicate substantial variation in epistemological reasoning even within the domains of social and natural science. For further corroboration of within-domain inconsistency, only 55% of respondents applied consistent epistemic stances to the two moral items. However, if there were no within-domain consistency, three items would be fully consistent only 11% of the time, six items only 4%. The consistency observed outstrips these chances considerably. Thus, the present evidence indicates a moderate degree of within-domain consistency.

Taking all eleven items as one scale (Absolutist=1, Multiplist=2, Evaluativist=3) produced a Cronbach's alpha of .74, within the >.70 range considered acceptable for multi-dimensional scales (Nunally, 1978). In an unrotated principal components analysis, all items except Nat 6 load (loading ≥ .39) on the first component. Nat 6 is the only explicit contradiction, "The universe will end" vs. "The universe will not end"; this may explain why it tends to elicit epistemic stance differently from the other items. The data also support some degree of cross-domain epistemological consistency. Epistemic stance in the natural science domain correlated with epistemic stance in the social science domain (r=.30, p<.001, for both three-item and six-item versions). The moral domain correlated weakly with the natural science domain (r=.17, p<.001) and slightly more with the social science domain (r=.24, p<.001). If we take each epistemic stance as its own scale, with items coded as 1 for each response consistent with that epistemic stance and 0 otherwise, then we can derive three scales: an Absolutism scale with an alpha of .69 (95% CI = .66, .72), a Multiplism scale with an alpha of .83 (95% CI = .82, .85), and an Evaluativism scale with an alpha of .82 (95% CI = .80, .84). Thus, there does seem to be substantial cross-domain consistency of epistemic stance, although not necessarily correlation of adjacent stages.

Developmental progression. Recall that the ESD model postulates a developmental sequence in order of increasing sophistication from absolutism to multiplism to evaluativism. This would suggest that a cohesive scale could be constructed by assigning values of 1, 2, and 3 respectively to the three epistemic stances for each item. Indeed, analyzing the eleven items in this way yields an acceptable Cronbach's alpha of .74, as mentioned above. However, if multiplism and evaluativism are switched in the sequence, the alpha doesn't change (alpha=.75). If we begin the sequence with multiplism instead of absolutism (multiplism=1, absolutism = 2, evaluativism = 3), the alpha actually rises to .85. In short, the sequence predicted by the ESD model does not fit with the data any better than other possible sequences. This fits with the abundant evidence undermining this sequence described in Section 1.3.

Demographic differences. Females were more likely to give multiplist responses than males (Levene's test shows unequal variances, F=8.9, p=.003; t-test without equality of variances assumed t=3.89 (793.6), p<.001, Cohen's d=.34) and less likely to give absolutist responses (Levene's test shows equal variances, F=3.1, p=.08; t=-3.46 (827), p=.001, Cohen's d=.33). Both effects held across all three domains (natural science, social science, and morality). There was no effect of gender on the frequency of evaluativist responses (Cohen's d=-.05).

A small portion of the sample attended a high-end private school, and the private school participants tended to give fewer absolutist responses (Cohen's d=.51) than public or charter school students, although the sample size (N=29 private school students) was too small for this difference to reach significance.

Discussion. As a theory of epistemological development through adolescence and early adulthood, the ESD model predicts that young adolescents should seldom exhibit an evaluativist

epistemic stance. After all, many adults do not. Yet in this sample of ninth graders, evaluativism was the commonest epistemic stance expressed, especially for scientific claims.

Were these particularly sophisticated adolescents? The large majority of participants attended the only public high school in a district ranked at the 46th percentile for math and the 47th for reading, putting them just below average for the country (Bush Institute, 2015). It is therefore unlikely that the participants in the current sample are more epistemologically sophisticated than the average American adolescent.

Instead, it seems that contemporary American adolescents incline somewhat towards evaluativism, at least on this particular measure. At a minimum, saying that two explanations for one phenomenon could both have some rightness but one could be more right indicates a capacity for nuance beyond simple absolutism.

The epistemic technologies account is further corroborated by the finding that inconsistencies in the activation of epistemic stance occur not only between domains (as discussed in Section 1.1), but also within domains. The weak to moderate degree of within-domain and cross-domain consistency are both consistent with the epistemic technologies account of epistemological reasoning. According to this model, objectivist and subjectivist epistemic technologies are both dormant in the thinker, drawn upon selectively for particular situations according to cues of topic and context. This gives rise to some consistency, because there are individual differences in the readiness-to-hand of particular epistemic technologies. It also gives rise to some inconsistency, because having an epistemic technology does not guarantee drawing upon it in a particular case, even when it would be appropriate. The more different two cases are, the less likely a thinker is to draw the same epistemic technologies from her repertoire. Thus,

cases from the same domain are more likely to be treated the same than cases from different domains.

Conclusions taken from this measure must be tentative, because the items may be interpreted in multiple ways. The original items oppose "different explanations" or merely "different books' explanations," without saying explicitly that these explanations conflict. It is pragmatically implied that they conflict, since the two fictional characters "believe" different books or people, but it is not explicitly stated and some participants may not assume they conflict at all. Moreover, there is a difference between opposing *explanations* and opposing *claims*; apparently opposing explanations frequently turn out to both hold some "rightness," as in the nature/nurture debate. Opposing claims are less often reconcilable. It is even possible that the failure of this measure to demonstrate the ESD model's postulated development trajectory is an artifact of the poorness of the measure, although even a very noisy measure should have picked up more development than this one did. Interviews are necessary to get a better grasp on how adolescents are conceptualizing the possible relationships between "different explanations." In any case, the paucity of evidence for the ESD model opens the door for the epistemic technologies model.

Another major point of dissent between the ESD model and the epistemic technologies model is the extent to which objective (absolutist) and subjective (multiplist) epistemic technologies will coexist before they are fully integrated. The epistemic technologies model claims that such coexistence will be very common even in young adolescents, while integration of the two occurs gradually through adolescence and adulthood. The ESD model claims that such coexistence will be less common, especially within a single domain. In order to investigate these claims, I turned to interviews with young adolescents on the nature of scientific disagreement and argument.

Study 2: Epistemology Interviews

Kuhn (1999, 2009) suggested that the recognition of the possibility of expert disagreement is key in sparking the transition from an absolutist epistemic stance to a multiplist one. She contends:

> The discovery that reasonable people—even experts—disagree serves as a source of recognizing the uncertain, subjective aspect of knowing. This recognition initially assumes such proportions, however, that it eclipses recognition of any objective standard that could serve as a basis for evaluating conflicting claims. From the multiplist perspective, knowledge consists not of facts but of opinions, freely chosen by their holders as personal possessions…Knowledge is now clearly seen as emanating from the knower, rather than the known, but at the significant cost of any discriminability among competing knowledge claims. (Kuhn, 2009, p.113).

Other thinkers agree that recognizing the possibility of expert disagreement is a major spur to recognizing that knowledge can be problematic, i.e. elusive and uncertain because underdetermined by the evidence (Carey & Smith, 1993). If even experts can disagree, then knowledge construction cannot be perfectly straightforward. Different ideas about what makes knowledge problematic will manifest as different sources of expert disagreement.

However, I am skeptical of Kuhn's contention that the recognition of expert disagreement initially "eclipses recognition of any objective standard that could serve as a basis for evaluating conflicting claims." Expert disagreement means that such an objective standard is not immediately and straightforwardly evaluable. However, it does not necessarily mean that there is no principled way to go about choosing between competing theories. Scientists are accustomed to both expert disagreement and principled epistemic strategies for evaluating competing theories, however tentatively.

According to the ESD model, the recognition that experts can disagree is the catalyst for the shift from absolutism to multiplism. The recognition of principled epistemic strategies despite uncertainty is the foundation of evaluativism. Thus, the simultaneous recognition of both expert

disagreement and principled epistemic strategies should come about only in an evaluativist stance.

Recall that, according to most ESD theorists, evaluativism should be fairly rare in adolescence (Kuhn, 1991, 2001, 2009; King & Kitchener, 1996). According to the epistemic technologies model, however, sources of expert disagreement and principled epistemic strategies are discrete epistemic technologies which can come online quite early in development, well before a fully cashed-out and integrated understanding of knowledge construction. The ESD model thus predicts that expert disagreement and epistemic strategies should coexist rarely in adolescents, while epistemic technologies theory predicts their coexistence should be common, although their integration might be gradual. In order to test these competing predictions, we turned to one-on-one interviews, again in the critical age of adolescence.

Procedure. Ninety-six participants at six schools were interviewed by trained interviewers mid-year in 8th grade, the year before the epistemic stance items were administered, and 84 of those were interviewed again 12 months later, in 9th grade. The portion of the interview discussed here was the last part of a longer interview, beginning after 20-40 minutes of questions about their goals, habits, and relationships. The epistemological portion of the interview focused on disagreement in natural science, specifically concerning the question of why earthquakes happen when and where they do.

Sample. The original 96 participants were 52% female, 46% African-American, 23% Caucasian, 16% Asian-American, 11% Hispanic, and 4% multi-racial. Nine attended a private school, 39 one of two public schools, and 48 one of four charter schools. Eighty-four were re-interviewed one year later; the remainder had left their original school district and could not be located.

Interview protocol: Disagreement in natural science. The interview probed student conceptions of knowledge of natural science and the nature of disagreement among scientists. Earthquakes were chosen as an interesting but little-known subject for most eighth grade students. Scientists know something about what makes earthquakes happen, but not everything is explained, leaving space for reasonable disagreement.

Interviewers first asked whether two expert scientists could ever disagree (querying the fallibility of expert knowledge):

> *A big part of science is explaining things. Seismologists are scientists who study earthquakes, and try to explain them. Do you think two seismologists will always agree on the best way to explain why earthquakes happen? Why/Why not?*

Next, they asked what would constitute a good scientific argument and how the student might go about figuring out which theory was right (i.e., what are good epistemic criteria?):

> *Let's suppose they do have two different theories. How would one scientist go about convincing the other that her theory was a better theory? How could you know whose theory was better?*

Finally, interviewers asked if the method given could give them the truth for sure, and if not, was there any method that could:

> *If you did that, could you be wrong? How could you find out?*

Students who said even the best methods would be fallible have at least one element of a more sophisticated epistemic stance than absolutism. Students who offered principled methods for choosing between theories could not be thoroughgoing multiplists. Any participant who acknowledged the possibility of expert disagreement, offered principled criteria for choosing between theories, and accepted the inevitable fallibility of even the best methods would exhibit the most essential elements of evaluativism (King and Kitchener, 1994; Kuhn, 1999, 2001).

Interview analysis. Each interview was coded according to a rubric of ideas or remarks which were either mentioned (coded 1) or not mentioned (coded 0). The first 20 interviews were coded by two coders; Cohen's kappas ranging from .61 to 1 were achieved, depending on the code, with a mean of .75 (Cohen, 1960). This puts all codes within the range of "substantial agreement" identified by Landis and Koch (1977) and the mean at the cusp between "good" and "excellent" agreement (Fleiss, 1981). General impression scores correlated at .85. The remaining interviews were coded by one coder. All coding was blind to school, demographics, and quantitative scores.

Results and Discussion.

Expert disagreement. Recall that all participants were asked whether expert scientists would always agree about why earthquakes happen when and where they do. Fully 87% (78) of the interview participants in eighth grade and 95% (73)[3] in ninth grade said that even expert scientists could disagree about the best explanation for earthquakes. Martha's[4] answer was typical: "Probably not, because people think in different ways." Test-retest reliability of this code indicated moderate stability over the twelve-month period (tetachoric correlation = .57, p=.02). While 2% (N=2) said expert disagreement was possible in eighth grade but regressed to denying it as ninth graders, 8% (N=7) switched in the other, predicted direction. Overall, recognition of expert disagreement was the norm, with only a few exceptions in either year.

Sources of expert disagreement. Many participants were able to give specific sources of expert disagreement, all reasonable possibilities. A few (7% (6) in 8th grade, 5% (4) in 9th grade) said the disagreement could arise from a mistake made by one of the scientists. For example, Daja said "There could have been an error in something they did." This is the only source of expert

[3] Some percentages are out of a smaller number than the total number of participants interviewed because some codes could not be evaluated due to interviewer error.
[4] All names are pseudonyms to preserve the anonymity of participants.

disagreement Kuhn says could be recognized even from an absolutist stance (Kuhn, 2009). However, three of the four ninth graders who cited the possibility of mistakes also mentioned another source of disagreement. According to the ESD model, mistakes are the easiest source of disagreement to recognize. Nonetheless, they are a genuine source of disagreement in science. In attributing scientific disagreement to error, these students are not drawing on subjectivist epistemic technologies. But they are not wrong.

Many participants pointed to additional legitimate sources of disagreement which indicated a dawning appreciation for the role of the knower in knowledge, and the need to go beyond mere "facts" (see Table 6). Nearly a quarter of participants had the insight that scientists might have different theories because they had different evidence (24% (23) in eighth grade; 17% (14) in 9th grade: test-retest tetrachoric correlation = .24, ns). Bella explained, "I think they could find information in different places." Others said the disagreement could arise from differences of interpretation from the same evidence, a crucial element of evaluativist reasoning (6% (5) in 8th grade, 11% (9) in 9th grade). Tyler said it well: "They won't always agree. It all depends on how you interpret the results of the research." Several[5] also mentioned the use of different methods, which presumably lead to different evidence ("different…techniques of doing things" or "different ways to solve the problem"). Most commonly, students suggested that the scientists might have different "points of view" or "different minds," mentioned by 40% (38) in eighth grade and 12% (10) in ninth grade.

[5] Different methods as a source of expert disagreement was too rare to obtain a reliable code.

Table 6

Expert Disagreement

	8th Grade	***9th Grade***
Experts Can Disagree	95% (78)	87% (73)
Because of…		
Mistakes	7% (6)	4% (4)
Different Evidence	24% (23)	17% (14)
Different Interpretations	6% (5)	11% (9)
Different Point of View	40% (38)	12% (10)

Only six of the 96 participants were so caught up in the inevitability of disagreement that objective standards were eclipsed, and two of these took it back upon reflection. These six are described in the next section. The vast majority of participants appealed to normative epistemic strategies for choosing between competing theories, described in later sections.

Expressions of multiplism. Six of the 96 participants (6%) did express skepticism about the possibility of any principled criteria or strategy for arbitrating between competing claims, two as eighth graders and four in ninth grade. None expressed multiplism about natural science in both years.

Consistent with Kuhn's (2009) theory that the recognition of expert disagreement would spur multiplism, all six of these participants first recognized the possibility of expert disagreement. For example, Maya expressed multiplism in terms very close to those of the epistemic stance items used in the quantitative measure (Kuhn et al., 2000):

> I don't think anybody's theory is better than the other one because everybody has different evidence. Everybody has different opinions and been through different things. So, nobody's theory is better than the other.

Maya's expression, "nobody's theory is better than the other," is very similar to that used in the Kuhn, Cheney, & Weinstock's (2000) quantitative measure used in Study 1, *One view could not be more right than the other.* Here there is not only no *method* to choose between theories—no principled epistemic strategy is possible—but also no *motivation* to do so; if all theories are equally right, it does not matter which one you hold. Maya does mention evidence. However, evidence is mentioned not as an epistemic strategy for reducing disagreement, but rather as a source of disagreement, "because everybody has different evidence."

Similarly, Joe gave a fairly clear expression of multiplism:

> Interviewer: How could you find out whose theory was the better theory?
>
> Joe: I'll look at one scientist's idea, then I'll look at the other, and see which one I like best.
>
> Interviewer: If you did that, could you be wrong about your choice?
>
> Joe: Well, for me, I wouldn't be wrong. But to someone else, there's different opinions…So somebody else may think I'm wrong 'cause that's their opinion.

Note that the epistemic strategy given — "see which one I like best" — fits with a multiplist epistemic stance, as does the relativism suggested by "*for me*, I wouldn't be wrong." Note also that this strong form of multiplism, as theorized by Kuhn (2009), undermines both the motivation and the possibility of argument based on shared principles, and consequently the motivation to search for better theories. If one is not wrong for oneself, but is wrong "to someone else," and both positions are equally valid, then there is no epistemic reason to try to understand the other's theory or to test one's own. Moreover, there is no principled way to choose between available epistemic strategies aside from preference, leaving Joe to choose whichever theory he "likes best."

However, the two participants who expressed multiplist intuitions in eighth grade took it back within the same interview. Take Jack:

> Jack: I don't know [if you could be wrong] because it's an opinion what's the main cause of earthquakes.
>
> Interviewer: What do you mean it's an opinion?
>
> Jack: It's not really an opinion. It's a fact that the natural way is the main cause of earthquakes.
>
> Interviewer: How could you be sure?
>
> Jack: Whoever had more evidence.

For a passing moment, Jack flirts with full-blown multiplism, but quickly slips back into an appeal to evidence.

Similarly, Milo expressed multiplism in eighth grade, but followed it up by asserting that there is a fact of the matter, after all, and one's feelings are not good guides to truth:

> Interviewer: How could you know whose theory was better?
>
> Milo: It depends what you believe about a theory. If you have a positive or a negative feeling towards a theory.
>
> Interviewer: So if you did that could you be wrong?
>
> Milo: It's a 50/50 chance you could be wrong or right.

These ambivalent participants certainly have some multiplist epistemic technologies—they appreciate the role of subjectivity in constructing beliefs. However, they do not therefore abandon the belief that some theories are better or truer than others. They do not have clear epistemic strategies for choosing between candidate theories, but they are not quite willing to take whatever theory they happen upon as true "for me."

The four ninth grade participants do seem to be genuinely ensconced in the multiplist stance. However, this 5% is a smaller proportion than the ESD model would predict, given the prevalence of recognizing expert disagreement (95% by ninth grade). Of those who recognized expert disagreement in ninth grade, only 4% denied either that one theory could be more right than the other or that there was any principled epistemic strategy for deciding. This fits with the

epistemic technologies model, in which elements of multiplism (recognition of subjective aspects of knowledge) can coexist with elements of absolutism (recognition of the objective aspects of knowledge) quite early in development.

Epistemic strategies in science. The large majority of participants (98% in 8th grade, 86% in 9th grade) did offer at least one principled epistemic strategy for choosing between theories. A grasp of and facility with effective epistemic strategies is at the core of epistemological sophistication in practice. For a full-blown multiplist, non-multiplist epistemic technologies should be rejected. The only legitimate epistemic strategies for a multiplist would involve exclusively subjective criteria; a matter of consulting one's own intuition or preferences, or perhaps consulting one's preferred personal authority like a parent or religious leader, with an admission that others could and even should consult their own personal authorities with very different results. Full-blown epistemic multiplism is an abdication of any attempt to agree across differences, an abdication of accountability to any criteria for belief shared beyond an immediate community chosen by personal preference or history. Insofar as students advocate or employ epistemic strategies that are not purely subjective, they are not thoroughgoing multiplists.

Recall that epistemic strategies for natural science were elicited in two ways: participants were first asked how one scientist would go about persuading the other scientist that her theory was better, and then asked how they themselves would go about choosing the better theory. The participants appealed to a number of epistemic strategies in answering these questions. I found that these strategies could be reliably categorized into four distinct forms: *appeal to authority*, *appeal to evidence*, *appeal to intelligibility*, and *appeal to social decision-making* processes such as consensus or discussion (see Table 7). Each of the four strategies came in more and less sophisticated versions, consistent with the claim that epistemic technologies are accessed early in

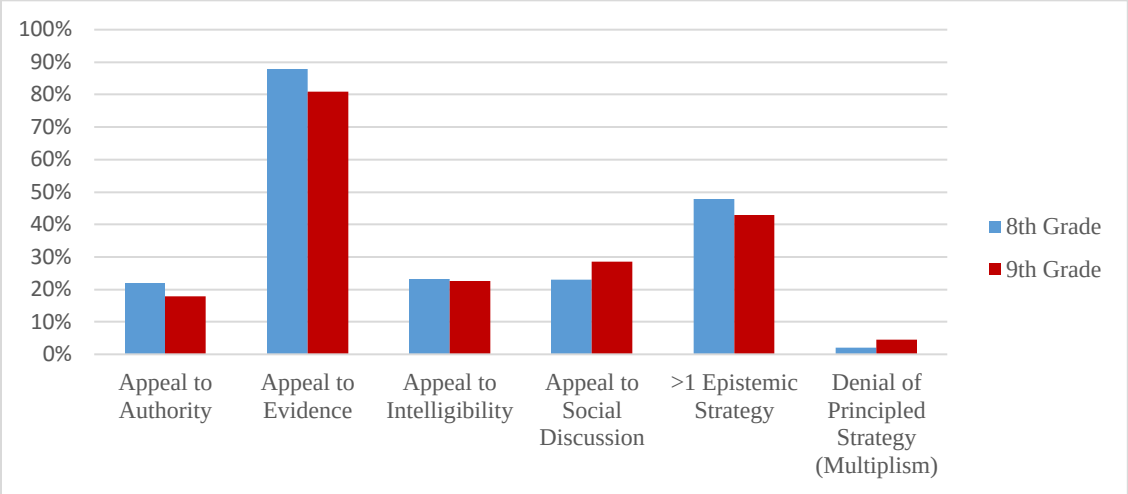
simple versions and gradually elaborated through experience and reflection. We will take up each of these epistemic strategies in turn.

Table 7

Overview of Epistemic Strategies

Epistemic Strategy	*Coding Schema*	*Example*	*Common Forms*
Appeal to Authority	Appeal to some higher authority (text or person) as method of theory selection	"Ask…a better scientist."	• Teacher • Internet/Text • Scientist
Appeal to Evidence	Appeal to evidence, data, facts, research, experiments, or proof as method of theory selection	"There's no way to know until you see the proper evidence."	• Proof • Facts • Experiment/research/ data • Evidence
Appeal to Intelligibility	Appeal to what makes sense, is logical, reasonable, plausible, or causally coherent as method of theory selection.	"Whichever one made more sense." "They're also able to use logic."	• "makes sense," can be understood • "logical" • "realistic," "plausible
Appeal to Social Discussion	Appeal to social discussion of theory and/or evidence as method of theory selection.	"I would ask other people's views on it, and see what they say."	• Detect mistakes • Pool data • Vote • Improve reasoning • Integration of theories

Figure 7: Frequencies of Epistemic Strategies



Appeal to authority. In eighth grade, 22% of the students cited appeal to authority as a good epistemic strategy, dropping insignificantly to 18% in ninth grade. Appeal to authority is a classic indicator of low epistemic sophistication, as listening to parents or a teacher does not require much understanding of how the authority acquired the knowledge or ability to evaluate their expertise for oneself.

However, professional scientists and epistemologists agree that appeal to the right authorities–often called *expert testimony*–is a necessary and legitimate form of epistemic justification for everyone, including experts. Knowledge is distributed, especially in the contemporary world; human knowledge is vast and complicated and experts tend to be highly specialized. Appeal to authority as a form of scientific argument is so common among scientists that it may be misleading to take it as an indicator of low sophistication among adolescents.

Consistent with this argument, none of the authorities cited by participants were unreasonable. The least sophisticated authorities were "Google," "my science teacher," "my notes from college," and "a textbook." Each of these was proposed by one student apiece, except for

looking online, which was cited by 8% (8) in 8th grade and 6% (5) in 9th grade. Any of these "authorities" plausibly do have more expertise than the students, though not more than the disagreeing scientists. Many appeals to authority were appeals to more expert scientists. For instance, Mari suggested, "Maybe like the head of the science department" could come and "straighten things out." Indicating more appreciation for the basis of expertise, Kira said, "I guess we would take it to a higher person that has been there for, that has been studying a lot of stuff for ages."

There are two keys to successful appeals to authority. First, the authority must be a legitimate authority on the subject. Even young children are quite good at tracking and inferring domains of expertise, suggesting this is something which may come naturally to us (Danovitch & Keil, 2004). A second and more challenging requirement for legitimate appeal to authority is that the appealer be justified in trusting their chosen authority. In many cases, perhaps most, we simply choose authorities on the basis of which authorities are cited or trusted by our interlocutors. However, this can sustain a pluralistic ignorance in which everyone in a community is citing a false authority, like climate change deniers all citing the two or three scientists who reject the evidence for climate change and ignoring the thousands of scientists who say otherwise. A more robust epistemic practice would require appeals to authority to be based in a meaningful appreciation for the reliability of the methods used by those authorities. Epistemic trust in authorities should be based on a justified appreciation for the epistemic effectiveness of their methods, not merely trust in their intentions.

The participants in this study appealed to authorities in two different ways, indicating very different levels of sophistication. Some appealed to authorities for a conclusive answer, as when Tia offered a straightforward way of finding out if she was right: "Ask my science teacher." Along the same lines, Bella said, "I could go and ask…a better scientist that actually knows the

reason why, not just the theories." Bella's response suggests that the original two scientists would disagree only because they weren't good enough scientists to know the answer. Others, however, appealed to authority in order to leverage other epistemic strategies, i.e., to improve their access to evidence or their ability to evaluate intelligibility. (The co-occurrence of appeal to authority with other epistemic strategies will be discussed later).

Appeal to evidence. The importance of appeal to evidence has received considerable attention in the science education literature. As such, it should not be surprising that this was the commonest form of scientific argument mentioned, cited by 88% (83) of students in eighth grade and 81% (68) in ninth grade. This widespread indication of respect for evidence casts doubt on the frequency of multiplism among adolescents indicated by the survey instrument, since a strict multiplism rejects the possibility of shared objective evidence.

The words used to refer to this strategy varied across individuals, but not grade. In eighth grade, 38% (36 participants) mentioned the need for "evidence," 53% (51) the need for "data," "research," or "experiment," 50% (48) the need for "proof," and 53% (51) the utility of "facts." None of these frequencies were significantly different a year later, when 43% (36) appealed to "evidence," 52% (44) mentioned "data," "research," or "experiment," 49% (41) "proof," and 49% (41) "facts." While "facts" are not quite the same thing as "evidence" for an experienced scientist, students treated them in the same way; both were forms of more or less objective information that could be accumulated from "research," possessed or not possessed, and shared with anyone interested. As such, appeal to 'facts' here counts as a form of appeal to evidence.

The least sophisticated version of appeal to evidence was an appeal to preponderance of evidence based on quantity rather than quality, variously characterized as, "Whoever has the most…evidence," "Whoever has better facts," or "The one that…has more data backing it up."

The emphasis on evidence indicated by its prevalence was also supported by the emphatic way participants spoke of it. As Jorge said, "If the other guy had more evidence than that one. It's all about evidence, if you don't got evidence you can't prove your claim." Other students said, "There's no way to know until you see the proper evidence," and "You have to have good evidence to back up every single claim you have." Note that none of these appeals to evidence—neither the appeals to quantity nor the insistence on its necessity—imply that all the facts will line up to support one theory. The epistemic strategy of going with "whoever has better facts" implies that both theories are consistent with *some* facts. A conception of knowledge as problematic is dawning here, even among the less sophisticated appeals to evidence.

Nonetheless, an explicit appreciation for the need for evidence may be only the first step towards proper use of evidence in scientific reasoning and argumentation. Sandoval (2003) found that trios of adolescents trying to explain a case of natural selection agreed about the need for evidence and did explore the available data in their quest for an explanation, but very few cited the relevant evidence to justify the explanations they turned up, despite ready access to a well-scaffolded platform for exploring a large dataset. Instead, they tended to rely on anecdotal evidence and what Sandoval terms *causal coherence*: an appeal to the intelligibility of the explanation as a causal mechanism for the phenomenon, independent of its fit with the data. This brings us to another common mode of justification mentioned: the appeal to intelligibility.

Appeal to intelligibility. Many students mentioned the relative intelligibility of a theory as a reason to accept it over alternatives, articulated variously as the one that "makes the most sense," or is "more understandable," "more likely to happen," "logical," or "more realistic." In eighth grade, this appeal showed up in 23% (19) of interviews; in ninth grade, in 16% (13), test-retest correlation *ns*. This was most frequently articulated as some form of what "makes more sense."

Intelligibility is a commonly observed criterion given by children and adolescents in the science education literature, although researchers tag it with many names. Samarapungavan (1992) found that even 7-year-olds could use logical consistency as a criterion of theory choice. In two papers from Sandoval (2003; Sandoval & Cam, 2011), children and adolescents both gave *evidentiary support* and *causal coherence* or *plausibility* as their top criteria for belief. The latter was usually referred to by students as "making sense," the commonest wording in the present interviews for what I coded as *appeal to intelligibility*.

The appeal to intelligibility does not necessarily indicate low sophistication. Those students in the present sample who elaborated enough to clarify indicated that causal coherence was the key to intelligibility. Derek, for instance, first said he would go with "whichever one was most logical. Which one makes the most sense." When the interviewer asked what he meant by this, he said, "Maybe because of how things are put together. Maybe what the cause is. If it seems realistic enough then I would believe that because it's realistic compared to something that seems less likely."

Several students explained what it meant to "sound logical" with an example of a plausible causal chain. Tia, for instance, said,

> Like a volcano, that's logical to me because earth's core is hot and everything, so there would be cracks coming up. I would say that's *logical,* and that's how volcanoes [could cause earthquakes].

Similarly, Jack, asked to elaborate his suggestion of picking whichever theory "makes more sense," offered, "They would say how more likely an earthquake would happen if...somebody was drilling in California for like, I don't know, gold, and they hit something, and it caused an earthquake."

The prevalence of both appeal to evidence and appeal to intelligibility is consistent with Sandoval's two studies. It also suggests that both of these criteria, commonly used by scientists, are already appreciated and utilized by young adolescents. This does not mean that they are using the criteria optimally. But it does suggest that students already understand, apply, and appreciate versions of normative criteria, which could be leveraged by teachers to develop more sophisticated and well-articulated versions of the same ideas. Conceptual development may be more in order than correction.

The undervaluing of the "making sense" criterion is more than a mistake of operationalization. Keeping track of what "makes sense" is at the core of genuine learning. In Posner, Strike, Hewson, and Gertzog's (1982) extremely influential paper on conceptual change, they suggested that two of the four keys for conceptual change are that the new conception must be (a) intelligible and (b) plausible. Many students are taught in science to accept ideas that are not intelligible to them, because memorization is often easier than wrapping one's mind around a complex and unfamiliar causal structure. However, memorization does not stick and cannot be transferred to new contexts. Learning to ignore the niggling feeling that something doesn't make sense is the antithesis of learning to think scientifically. Students may benefit from further articulating what exactly they mean when they say something doesn't "make sense," but this development can build on the original intuition. Indeed, some students said they might be better able to figure out what really "makes sense" on the basis of social discussion.

Appeal to social discussion. The fourth common epistemic strategy was to discuss the competing theories with multiple people, advocated by 22% (18) of eighth graders and 26% (22) of ninth graders. They offered a variety of justifications for this appeal to social discussion. Some justifications were rooted in objective criteria, i.e. (a) more people are more likely to detect mistakes, or (b) multiple researchers can pool their data so they have more evidence. Dora, for

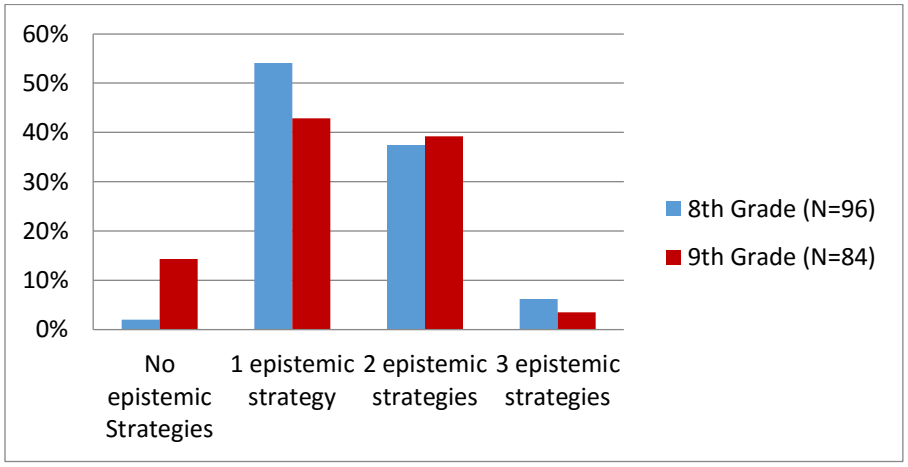
example, seemed optimistic that social discussion would help catch mistakes, saying, "It's like double checking and triple checking, and it's just, most people think this person is right, so it's probably right." Although expert disagreement through error is fully consistent with absolutism, as discussed earlier, the correction of errors is an invaluable function of collaboration and peer review in well-functioning scientific practice.

Others said they would base their choice on majority vote. For instance, Moira said, "I guess I would choose the side that more people agreed with." Similarly, Ali said she would choose a theory "based on majority pick." This could fit into a more multiplist epistemic stance, since if you think no answer is better than another but have to pick one, you might as well maximize the satisfaction of the community. However, Ali, at least, seemed to be appealing to "majority pick" in more objectivist or even evaluativist hopes, as she continued, "that one might be *more correct*" [emphasis added]. If one takes all members of a community to have some reasonable but diverse ways of approaching the best theory, then majority pick is not the worst strategy. Ali also uses the modal *might,* indicating she does not take it for an infallible strategy.

Most appeals to social discussion were rooted in integrated objective and subjective criteria, i.e. when the goal is to discuss to consensus based on other epistemic criteria. Discussion to consensus can include all of the previous forms: sharing data, detecting mistakes, and improving the quality of reasoning, and in addition integrating the best aspects of each theory and understanding why the evidence might not be sufficient to distinguish between competing theories. With so many functions, it is an excellent example of an epistemic technology which comes in both very intuitive and very sophisticated versions.

All but two of the 96 eighth graders (2%) and 12 of the 84 ninth graders (14%) offered at least one of these four epistemic strategies (see Figure 8). In both eighth and ninth grade, 44% (42 and 38) cited two or more of the strategies (test-retest correlation = .23, p=.03). (See Figure 8.)

Figure 8: Number of Epistemic Strategies per Participant



Coordination of epistemic strategies. The best way to choose the better of two candidate scientific theories is to coordinate multiple epistemic strategies. They showed varying levels of coordination of strategies, discussed in the sections below.

Coordination of appeals to evidence and intelligibility. Some participants connected the appeal to evidence with an appeal to intelligibility, noting that evidence requires intelligible interpretation to link it with competing theories. "Reason" is needed to see if the evidence "connects" to the theory in the right way. For instance, Brandon insists that no number of "facts" will be sufficient if they are not intelligibly connected to the theory:

> Showing me a lot of facts, and for me to comprehend and understand what you mean about those facts, not just like sitting down. You can have 1,000 facts, and I won't understand a word, which one you mean. But the other person have like two facts,

> and I'll understand everything, and I'll believe this one [theory] more than that one, not just by sight.

Similarly, Tyler says: "It would have to tie in more with the facts. So whatever makes more sense would be the more superior theory." The link between the theory and the facts must be determined to establish "what makes more sense."

Arya was even more explicit about the need for interpretation of the evidence and the possibility of conflicting interpretations:

> It's based on facts. And yes, so it's based on which fact is stronger. But at the same time, they both think their fact is stronger. That's why you need to have a different eye to look at it, maybe go ask somebody else what they think is more correct than the other person, the one that have more understanding, the one that generally seems more realistic than the other one.

She first cites the *preponderance of evidence* criterion discussed earlier, then admits that this cannot always be evaluated directly from the evidence, "They both think their fact is stronger." She then calls for a third party not merely to break the tie, but to add their reasoning to the discussion, explaining which one seems "more correct," and has "more understanding" (i.e., is intelligible) to them. Although she knows that reasonable experts can disagree about what the evidence means, she remains optimistic that the proper application of reasoning about the evidence will point to the better, "more realistic" theory, especially in a context of back-and-forth discussion between multiple experts.

Contrast these cases to Ryan, who said, "They're also able to use logic, and give them facts about how the reason why it's gonna happen, or why the activity will happen." Like Brandon, Tyler, and Arya, Ryan cites both appeal to evidence ("facts") and appeal to intelligibility ("how the reason why it's gonna happen"). Both epistemic strategies are in each participant's repertoire of epistemic technologies, and Ryan is the only one who does not integrate them.

Appeal to authority to enhance evidence and intelligibility. Many of those who appealed to authority did so not simply for a conclusive answer, but rather in order to improve the quality of other epistemic strategies. Specifically, some appealed to authority for more evidence (leveraging the epistemic strategy of *appeal to evidence*) and some for superior reasoning (leveraging the epistemic strategy of *appeal to intelligibility*). Ryan, for instance, used the phrase, "Appeal to authority," which he expanded upon by saying, "People will mainly trust someone who has studied in that area more…And they're also able to use *logic*, and give them *facts* about how the *reason why* it's gonna happen" (ital. mine). He is justifying the appeal to authority by pointing to their superior ability to utilize the epistemic strategies *appeal to intelligibility* ("they're able to use logic, and give… the reason why it's gonna happen) and *appeal to evidence* ("they're able to…give them facts"). This is the most epistemically responsible way to choose epistemic authorities: on the basis of their expertise in utilizing epistemic strategies we believe are reliable.

Social discussion to enhance evidence, intelligibility, and authority. As with appeal to authority, appeal to social discussion was often proposed as a way to augment the power of appeals to evidence and intelligibility. Indeed, the most widespread justification given for appeal to social discussion was to augment information/evidence and/or reasoning about what would "make sense." For instance, Hera said,

> I would ask other people's views on it, and seeing what they say about it…[That way] I would have more of a view of what makes sense to other people. And if it seems like the other theory makes more sense, then I'll look at it, and I'll try to figure out why they think it makes sense.

Similarly, Hugo said, "They'd have to look for flaws in each other's reasoning." For Hugo, the collaborative search for flaws of reasoning would improve the effectiveness of appeals to intelligibility.

Meanwhile, Leila focused on the utility of social discussion for pooling evidence and the resultant expertise (i.e., authority):

> Leila: Maybe if there were other scientists to back her up on it, or there was proof of her theory.
>
> Interviewer: Why would multiple scientists help?
>
> Leila: Because if they're scientists, so they do a lot of research and experiments. So if multiple scientists are backing her up on it, there's a really good possibility that she's right.

Note that here the *appeal to social discussion* epistemic strategy is fully integrated with the strategies of *appeal to authority* ("if they're scientists") and *appeal to evidence* ("they do a lot of research and experiments"). This makes sense in the epistemic technologies model; Leila is drawing upon three epistemic technologies in tandem, integrating them appropriately to create a more sophisticated and effective epistemic strategy. She would not be able to do this if she did not have all three epistemic strategies ready at hand. As further indication of sophistication, even the successful integration of all three of these strategies is given only a probabilistic confidence: "If multiple scientists are backing her up on it, there's a really good possibility that she's right."

It is reasonable that the integration of multiple epistemic strategies should come with increased appreciation for the uncertainty of resulting conclusions, since an appeal to multiple epistemic strategies can generate conflicting views. For instance, it is possible that other scientists could agree despite a lack of sufficient "proof," which would pit authority against evidence. Moreover, the social decision-making process itself makes the possibility of disagreement and hence uncertainty more salient.

Another common mode of appeal to social decision-making came in the form of a desire to integrate the competing theories, with the scientists collaborating to combine the best aspects of each theory. For example, Todd said, "He would probably think of a way to make his and the

other seismologist's theories work together to make one way." Giving a few more details, Ahmed said,

> They won't always come up with the same ideas, but they can combine those ideas to make it the perfect example why this is happening and find the reasoning between each other's faults and their explanation on what should be included and not included.

It is only worth putting together the best parts of each theory if both theories "have some rightness," to recall the phrasing of the quantitative measure (Kuhn, Cheney, and Weinstock, 2000), but neither holds the whole story. "They won't always come up with the same ideas" indicates an appreciation for the role of the subjective in knowledge construction. However, Ahmed thinks that through social discussion about the candidate theories, the reasoning and explanations, it may be possible to come to agreement on what elements of each view "should be included and not included."

Synopsis of Interviews. The ideal method of science is the integration of all four epistemic strategies, social discussion between experts appealing to both evidence and intelligibility, pooling data and ideas and catching each other's mistakes of evidence and reasoning. None of the adolescent participants described such a complete integration, nor did they exhibit fully sophisticated versions of the strategies. They still had much to learn to attain the full epistemological understanding of a trained scientist. However, they did draw upon many elements of such an understanding. For instance, the majority of adolescents demonstrated some grasp of both objective and subjective epistemic technologies (80% in eighth grade and 82% in ninth grade), recognizing some role of subjectivity in knowledge construction in their recognition of expert disagreement, and recognizing the role of objectivity in the form of appeals to evidence or facts. Moreover, all but two eighth graders were able to describe at least one principled epistemic

strategy for choosing the better theory, while conscious that even that would not yield a perfectly certain or perfect conclusion.

Contemplated from the view of the epistemic technologies model, the co-existence of subjective and objective epistemic technologies is what we would expect. Moreover, the varied levels of integration of epistemic strategies suggests that a major variable in individual differences is facility with and integration of the epistemic technologies, not merely their presence or absence in individuals' repertoires. Although very few individuals proposed all four epistemic strategies, I suspect that the right questions could have elicited all four from almost all students. Participants varied in the salience of particular epistemic technologies and in facility of their application, leading them to draw out different technologies in response to the same question. The low test-retest correlations support the idea that most students did have most of the relevant epistemic technologies in their repertoire, but drew out only enough to satisfy the interviewer. More probing interviews on a wider range of subjects would be needed to test this hypothesis.

The remarkably small difference between eighth and ninth grade was not predicted, but nonetheless fits the epistemic technologies model. If epistemological development is better characterized as the increasing sophistication and coordination of epistemic technologies, rather than a progression through distinct sets of epistemic technologies each constituting a stance, then we should not expect much change in the frequency with which epistemic technologies emerge over the course of a year. This is true even in a year crucial for cognitive development. However, there was no significant sign of increased coordination of epistemic technologies over the course of this year, which would have served as stronger evidence for the epistemic technologies model. Considerable further research is needed to test the utility of this model for educational practice and prediction.

Conclusion

I have argued that lay epistemology develops largely piecewise. That is, emerging thinkers begin with simplified versions of normative epistemic technologies at a young age, and gradually elaborate, integrate, and gain facility in using these technologies over the course of epistemological development. This contrasts with the currently dominant account of epistemological development defended by Kuhn (1991, 2002, 2009) and King and Kitchener (1994, 2003), which suggests that epistemological development occurs holistically, with a coherent and consistent epistemology which changes over development. By the principle of Henrion's Glue[6], there is most likely some truth in both models.

According to the ESD model, epistemological development begins with respect for the objective world and optimism about our ability to know it directly, graduates through the frustration of epistemic failures to an exclusive appreciation of the subjective, and finally, for some, achieves recognition of the epistemic pressures of both an objective external reality and subjective internal representations. The ESD model's insight into the tension between objective and subjective aspects of knowledge-building is crucial. However, I have argued that recognition of the objective is rarely overcome entirely by the subjective; instead, both coexist from a young age, and only gradually become integrated.

The ESD model thus (a) oversimplifies the coherence of epistemological thinking within individuals at a given time-point, and (b) underestimates the capacities of young children to appreciate and utilize elements of sophisticated epistemology. In this paper, I have presented evidence from both previously published and original research that the ESD model does not adequately account for the evidence. Instead, I suggest that individuals draw from a growing

[6] The Principle of Henrion's Glue states that when there two competing explanations for a phenomenon in psychology or sociology, chances are they both have some claim on the truth. For example, take the Nature/Nurture debate.

repertoire of epistemic technologies, including epistemic strategies like "appeal to authority," epistemic forms like "outline," or representational modes like "doubt." More sophisticated thinkers have a larger repertoire of more elaborated and coordinated epistemic technologies.

A better understanding of epistemological development can illuminate leverage points to scaffold and stimulate education in epistemology. This may be particularly important given the low level of epistemological understanding among most adults (Baxter Magolda, 2002; Kitchener, 1983; King, Kitchener, & Wood, 1994; Perry, 1970). Moreover, poor understanding of the nature of knowledge, especially scientific knowledge, is a predictor of poorer reasoning in both scientific and legal contexts (Lombrozo, Weisberg, & Thanukos, 2008; Weinstock, 2005).

If the epistemic technologies model holds water, what does that imply for educators? It suggests that educators should seek to leverage epistemic technologies students already have. For instance, teachers could give students opportunities to employ all four epistemic strategies described here to answer questions, one at a time and in coordination. Students could conduct science projects which integrate appeals to evidence, intelligibility, authority, and social discussion within the classroom. It might also be helpful to encourage students to make explicit the epistemic strategies they are using. Classroom discussions could be held to discuss criteria for what makes each epistemic strategy most effective: What counts as good evidence? Who should we take as a legitimate epistemic authority on this topic, and on what grounds? When we say a theory "makes sense," what do we mean? When we say a theory "doesn't make sense," to what problem are we pointing? What do we learn from classroom discussions like this?

My contention is not that American adolescents already understand everything they need to understand about knowledge; on the contrary, a sophisticated epistemology is rare even among college graduates. My contention is that even young adolescents have the basic elements of a

sophisticated epistemology. We need to meet students where they are and help them build upon the epistemic technologies they already have. Much of the literature on science education and personal epistemology emphasizes the ways people go wrong. And indeed they do go wrong all over the place. But they also have a lot right, even in early adolescence. The adolescents in the interviews described above were not wrong about the importance of intelligibility, or appeals to authority, or the need for social discussion of theories and evidence, or the multiplicity of existing opinions. We need teachers to help students put these pieces together in the right way, integrating the subjective and objective dimensions of knowledge construction to develop patterns of reasoning that are responsive to reality while acknowledging the underdetermination of theory by the evidence. Only from this integrated epistemological perspective can they respect both the power and the limitations of human knowledge.

Bibliography

Baxter Magolda, M. B. (1992). *Knowing and reasoning in college: Gender-related patterns in students' intellectual development.* San Francisco: Jossey-Bass.

Baxter Magolda, M. B. (2001). *Making their own way: Narratives for transforming higher education to promote self-development*. Sterling, VA: Stylus.

Baxter Magolda, M. B. (2004). Evolution of a constructivist conceptualization of epistemological reflection. *Educational Psychologist, 39:1,* 31-42.

Beebe, A., & Sackris, D. (*in press*). Moral objectivism across the lifespan. *Philosophical Psychology*.

Boyes, M. C. & Chandler, M. (1992). Cognitive development, epistemic doubt, and identity formation in adolescence. *Journal of Youth and Adolescence, 21,* 737-763.

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology, 20 (4),* 493-523.

Bush Institute (2015). Global Report Card. http://www.globalreportcard.org/

Carey, S. & Smith, C. (1993). On understanding the nature of scientific knowledge. *Educational Psychologist, 28, 3,* 235-251.

Carpendale, J. I., & Chandler, M. J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development, 67*(4), 1686-1706.

Cho, Y. H., Lee, J., & Jonassen, D. H. (2011). The role of tasks and epistemological beliefs in online peer questioning. *Computers & Education*, *56*(1), 112-126.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46.

Collins, A., & Ferguson, W. (1993). Epistemic forms and epistemic games: Structures and strategies to guide inquiry. *Educational psychologist*, *28*(1), 25-42.

Danovitch, J.H. & Keil, F. C. (2004). Should you ask a fisherman or a biologist? Developmental shifts in ways of clustering knowledge. *Child Development, 75, 3,* 918-931.

Detterman, D. K. (1993). The case for the prosecution: Transfer as an epiphenomenon.

Elby, A., & Hammer, D. (2010). Epistemological resources and framing: A cognitive framework for helping teachers interpret and respond to their students' epistemologies. *Personal epistemology in the classroom: Theory, research, and implications for practice*, 409-434.

Flavell, J. H., Flavell, E. R., Green, F. L., & Moses, L. J. (1990). Young children's understanding of fact beliefs versus value beliefs. *Child development*, *61*(4), 915-928.

Flavell J.H., Green F.L., Flavell E.R. (1990). *Cognitive Development, 5, 1.*

Fleiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability studies. *Applied Psychological Measurement*, *5*(1), 105-112.

Gick, M. L. & Holyoak, K. J. (1983) Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1-38.

Hammer, D., & Elby, A. (2002). On the form of a personal epistemology. *Personal epistemology: The psychology of beliefs about knowledge and knowing*, 169-190.

Hammer, D., & Elby, A. (2003). Tapping epistemological resources for learning physics. *The Journal of the Learning Sciences*, *12*(1), 53-90.

Elby, A., & Hammer, D. (2010). Epistemological resources and framing: A cognitive framework for helping teachers interpret and respond to their students' epistemologies. *Personal epistemology in the classroom: Theory, research, and implications for practice*, 409-434.

Hofer, B. K. (2006). Domain specificity of personal epistemology: Resolved questions, persistent issues, new models. *International Journal of Educational Research*, *45*(1), 85-95.

Hofer, B. K. (2006). Beliefs about knowledge and knowing: Integrating domain specificity and domain generality: A response to Muis, Bendixen, and Haerle (2006). *Educational Psychology Review, 18, 1,* 67-76.

Holyoak, K.J. (2012). Analogy and relational reasoning. In K.J. Holyoak & R.G. Morrison (Eds.), The Oxford Handbook of Thinking and Reasoning (pp. 234-259). New York: Oxford University Press.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.

Louca, L., Elby, A., Hammer, D., & Kagey, T. (2004). Epistemological resources: Applying a new epistemological framework to science instruction. *Educational Psychologist*, *39*(1), 57-68.

Luce, M.R.; Callanan, M.A.; Smilovic, S. (2012). Parents' epistemological stance and children's evidence talk. *Developmental Psychology, 49, 3,* 454-461.

King, P. M., & Kitchener, K. S. (2002). The reflective judgment model: Twenty years of epistemic cognition. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 37–61). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

King, P. M., Kitchener, K. S., & Wood, P. K. (1994). Research on the reflective judgment model. *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults.* San Francisco: Jossey-Bass.

Kitchener, K. S. (1983). Cognition, metacognition and epistemic cognition: A three-level model of cognitive processing. *Human Development, 4,* 222–232.

Krettenauer, T. (2004). Metaethical cognition and epistemic reasoning development in adolescence. *International Journal of Behavioral Development*, *28*(5), 461-470.

Krettenauer, T. (2005). The role of epistemic cognition in adolescent identity formation: Further evidence. *Journal of Youth and Adolescence*, *34*(3), 185-198.

Kuhn, D. (1991). *The Skills of Argument.* New York: Cambridge University Press.

Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher, 28, 2,* 16-26, 46.

Kuhn, D. (2001). How do people know? *Psychological Science, 12, 1,* 1-8.

Kuhn, D. (2009). The importance of learning about knowing: Creating a foundation for development of intellectual values. *Child Development Perspectives*, *3*(2), 112-117.

Kuhn, D.; Cheney, R; & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development, 15,* 309-328.

Kuhn, D., & Park, S. H. (2005). Epistemological understanding and the development of intellectual values. *International Journal of Educational Research*, *43*(3), 111-124.

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 363-374.

Louca, L., Elby, A., Hammer, D., & Kagey, T. (2004). Epistemological resources: Applying a new epistemological framework to science instruction. *Educational Psychologist, 39*(1), 57-68.

Mansfield, A. F. & Clinchy, B. M. (2002). Toward the integration of objectivity and subjectivity: Epistemological development from 10 to 16. *New Ideas in Psychology, 20,* 225-262.

Mason, L., Boldrin, A., & Zurlo, G. (2006). Epistemological understanding in different judgment domains: Relationships with gender, grade level, and curriculum. *International Journal of Educational Research, 45*(1), 43-56.

Mason, L.; Boscolo, P.; Tornatora, M.C.; & Ronconi, L. (2013). Besides knowledge: A cross-sectional study on the relations between epistemic beliefs, achievement goals, self-beliefs, and achievement in science. *Instructional Science, 41,* 49-79.

Mason, L., & Scirica, F. (2006). Prediction of students' argumentation skills about controversial topics by epistemological understanding. *Learning and Instruction, 16*(5), 492-509.

Metz, K.E. (2004). Children's understanding of scientific inquiry: Their conceptualization of uncertainty in investigations of their own design. *Cognition and Instruction, 22, 2,* 219-290.

Moschner, B.; Anscheutz, A.; Wernke, S.; & Wagener, U. (2008). Measurement of epistemological beliefs and learning strategies of elementary school children. In *Knowing, Knowledge, and Beliefs: Epistemological Studies across Diverse Cultures* (Ed. Khine, M.S.). pp.113-133. Springer.

Moshman, D. (2015). *Epistemic cognition and development: The psychology of justification and truth*. Psychology Press.

Muis, K. R.; Bendixen, L.D.; & Haerle, F.L. (2006). Domain-generality and domain-specificity in personal epistemology research: Philosophical and empirical reflections in the development of a theoretical framework. *Educational Psychological Review, 18,* 3-54.

Nichols, S. (2004). After objectivity: An empirical study of moral judgment. *Philosophical Psychology, 17,* 3-26.

Novick, L. R. (1990). Representational transfer in problem solving. *Psychological Science, 1*(2), 128-132.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Perry, W (1970). *Forms of Intellectual and Ethical Development in the College Years.* Holt, Rinehart, & Winston, Inc.

Posner, G.J.; Strike, K. A.; Hewson, P. W.; & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66(2),* 211-227.

Olafson, L., Schraw, G., & Vander Veldt, M. (2010). Consistency and development of teachers' epistemological and ontological world views. *Learning Environments Research*, *13*(3), 243-266.

Reed, S. K. (1987). A structure-mapping model for word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 124-139.

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: evidence from 14-and 18-month-olds. *Developmental psychology*, *33*(1), 12.

Ricco, R.; Pierce, S. S.; & Medinilla, C. (2010). Epistemic beliefs and achievement motivation in early adolescence. *Journal of Early Adolescence, 30, 2,* 305-340.

Samarapungavan, A. (1992). Children's judgments in theory choice tasks: Scientific rationality in childhood. *Cognition, 45,* 1-32.

Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *The Journal of the Learning Sciences, 12(1)*, 5-51.

Sandoval, W. A. & Cam, A. (2011). Elementary children's judgments of the epistemic status of sources of justification. *Science Education, 95,* 383-408.

Sarkissian, H., Park, J., Tien, D., Wright, J. C., & Knobe, J. (2011). Folk moral relativism. *Mind and Language, 26,* 482-505.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 0956797611417632.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: a key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534.

Stathopoulou, C.; & Vosniadou, S. (2007). Exploring the relationship between physics-related epistemological beliefs and physics understanding. *Contemporary Educational Psychology, 32,* 255-281.

Sousa, P., Piazza, J., & Goodwin, G. P. (2016). *Folk moral objectivism*. Manuscript in preparation.

Tsai, C.C.; Ho, H. N. J.; Liang, J.C.; Lin, H.M. (2011). Scientific epistemic beliefs, conceptions of learning science and self-efficacy of learning science among high school students. *Learning and Instruction, 21,* 757-769.

Wainryb, C.; Shaw, L. A.; Langley, M.; Cottam, K.; & Lewin, R. (2004). Thinking about diversity of belief in the early school years: Judgments of relativism, tolerance, and disagreeing persons. *Child Development, 75 (3),* 687-703.

Weinstein, M. & Cronin, M. A. (2003). The everyday production of knowledge: Individual differences in epistemological understanding and juror-reasoning skill. *Applied Cognitive Psychology, 17,* 161-181.

Weinstock, M. P. (2005). Cognitive bases for effective participation in democratic institutions: Argument skill and juror reasoning. *Theory & Research in Social Education*, *33*(1), 73-102.

Weinstock, M. P., Neuman, Y., & Glassner, A. (2006). Identification of informal reasoning fallacies as a function of epistemological level, grade level, and cognitive ability. *Journal of Educational Psychology*, *98*(2), 327.

Wellman HM, Bartsch K. (1988). Young children's reasoning about beliefs. *Cognition, 30 (3),* 239-277.

Wimmer H, & Perner J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13 (1),* 103-128.

CHAPTER 3

NORMS AND HABITS OF ACTIVELY OPEN-MINDED THINKING IN AMERICAN ADOLESCENCE

Abstract

In a mixed-methods 18-month study of 1551 adolescents from 8 diverse schools across the U.S., the large majority of adolescents demonstrated (a) strong norms of actively open-minded thinking and (b) a widespread capacity for actively open-minded thinking. A rating scale, two new multiple choice measures, teacher report, peer nomination, and 96 interviews converged to reveal widespread openness to disagreement, desire for understanding, and pluralistic norms. In interviews, participants demonstrated actively open-minded thinking in three ways: a capacity for *deep search for ideas*, *epistemic empathy*, and *pluralistic thinking*. Possible explanations of adolescents' deep capacity for actively open-minded thinking are discussed.

Introduction

Ever since Dewey (1916), there has been a growing push to teach students not just *what* to think, but *how* to think. Eighty percent of Americans think schools should teach critical thinking (Gallup, 2013). Critical thinking is front and center in the Common Core Standards, a set of curricular goalposts rolled out in 2009 and now adopted by 42 states (Common Core State Standards Initiative, 2016). But if we as a society want to teach our children to be good thinkers, we had better have some idea how a good thinker thinks.

Baron (1988, 1993) suggested that the best critical thinking is *actively open-minded;* that is, characterized by deep and unbiased search for alternatives, evidence, reasons, and goals. Long reflection is not sufficient for good thinking if it is merely an elaboration of support for a favored conclusion (Wason & Evans, 1975); a balanced consideration of alternatives is also necessary.

Most cases of poor decision-making arise from a search that is either not deep enough (i.e. too shallow) or too biased. Problems of shallow search include inadequate information search and failure to consider alternatives at all (Cacioppo & Petty, 1982; Cacioppo, Petty, Feinstein, &

Jarvis, 1996; Runco, 2010). Biases include the ubiquitous confirmation bias (Nickerson, 1990) and many other cases of motivated reasoning reviewed by Kunda (1990), from ego-protective processing of data (e.g. Wyer & Frey, 1983; Kassarjian & Cohen, 1965; Kunda, 1987) to goal-biased statistical reasoning (e.g. Ginossar & Trope, 1987). Actively open-minded thinking is thus equally valuable for decision-making about what to do and judgments of what to believe. It is thus theoretically plausible to take actively open-minded thinking (AOT) as a centerpiece of good critical thinking.

Empirical evidence supports this thesis. Even self-reported preferences for AOT can predict the ability to solve many critical thinking tasks. Stanovich and colleagues have found that among adults, AOT predicts superior evaluations of the quality of arguments, superior performance on counterfactual syllogisms and other heuristics and biases, and less motivated reasoning in argument evaluation (Stanovich & West, 1997, 2007, 2008; West, Toplak, & Stanovich, 2008). In combination with other closely related scales, it also predicts better statistical reasoning and covariation detection, and less outcome bias and if-only thinking (Stanovich & West, 1998). AOT is similarly important among children, predicting better performance on counterfactual syllogisms, base rate sensitivity, and sensitivity to framing (Kokis, Macpherson, Toplak, West, & Stanovich, 2002). Despite the exclusive use of self-report scales to measure AOT, almost all of these effects in both adults and children are independent of other ability measures.

Stanovich and West (1998) describe AOT in terms of four facets: "tendency to consider alternative opinions and evidence," "willingness to decontextualize," "willingness to switch perspectives," and "epistemological non-absolutism." The tendency to consider alternative opinions and evidence is what we are calling here *deep search for possibilities and evidence*. Willingness to decontextualize and willingness to switch perspectives can be seen as two faces of

one coin. Switching perspectives requires replacing one's own present context with someone else's context, and decontextualizing requires noticing which aspects of one's context are not shared by other perspectives. I propose that we call the capacity to engage in such context- and perspective-switching *epistemic empathy,* the ability to understand the intellectual position and internal logic of someone with whom one disagrees. Epistemological non-absolutism—or, to put it more normatively, a certain degree of *epistemological sophistication*—is necessary for an individual to believe they could learn something new from actively open-minded thinking.

I have explored epistemological non-absolutism in depth in Chapter 2 of this dissertation. In the current chapter, I will explore the other key facets of actively open-minded thinking: deep search for possibilities and evidence and epistemic empathy.

The need for deep search for possibilities and evidence. Much of the literature on cognitive style has examined situational and individual differences in the disposition to search deeply for possibilities and evidence. Two of the most widely used cognitive style measures are *need for cognition* and *need for closure*. *Need for cognition* is the disposition to enjoy cognitively effortful activities, which often manifests as a willingness to engage in deep search for possibilities and evidence (Cacioppo & Petty, 1982; Cacioppo, Petty, & Kao, 1986). Need for cognition predicts better problem-solving and verbal reasoning, higher tolerance of ambiguity and uncertainty, higher GPA, and more complex social attributions (see Cacioppo, Petty, Feinstein, & Jarvis, 1996 for an extensive review).

Conversely, *need for closure* is the tendency to freeze upon one position and avoid alternatives or evidence against it (Webster & Kruglanski, 1994). Individuals with a high need for closure seem to dislike even temporary periods of uncertainty when they must try to find an answer. Need for closure predicts a greater tendency to stereotype (Dijksterhuis, Knippenberg,

Kruglanski, & Schaper, 1996), greater conformity and less egalitarian discussion (De Grada, Kruglanski, Mannetti, Pierro, 1999), and a preference for homogeneous groups (Kruglanski, Shah, Pierro, & Manetti, 2002). Both need for closure and need for cognition are usually measured with self-report rating scales.

Actively open-minded thinking requires a willingness to put in the effort for deep search and postpone closure, sometimes indefinitely. The wide variety of valuable correlates described offer empirical support for the thesis that actively open-minded thinking could serve as an effective centerpiece of good critical thinking.

The need for epistemic empathy. *Epistemic empathy* is understanding what it is like to believe what another person believes, complete with their internal reasons for that belief and its implications. This does not require supposing the other person is *correct,* or even changing one's own beliefs at all. The literature on cognitive perspective-taking is focused on the ability to understand that someone else might have different beliefs at all. Epistemic empathy requires more than this; it requires one not only to understand what others believe, but also why.

Understanding another's position may reveal unexpected virtues of that position—or it may not. Regardless, it is impossible to adequately judge a position without understanding its place within its own framework of logic and values. Thus, a capacity for epistemic empathy is a great support for intellectual development, enabling far more productive conversations across diverse perspectives. Without epistemic empathy, dialogues between people from different backgrounds tend to be ships passing in the night, both trying to communicate with flashing lights, one using Morse and the other using 5x5 tap code, each increasingly frustrated with the incoherence of the other's flashes.

Epistemic empathy is both the stimulus and the result of an actively open mind. In showing why others might think differently, epistemic empathy forces an examination of alternative possibilities and evidence or reasons against our own beliefs. Epistemic empathy may even be the mainspring of human capacity for actively open-minded thinking. Vygotsky (1986) and Yegnashankaran (2010) have argued that internal reasoning originates in external dialogue with others, where children learn how reasons are demanded and exchanged. Insofar as this is true, those who cannot understand why someone might disagree with them will not consider alternative perspectives or possibilities. Epistemic empathy for the disparate beliefs of those around us is the basis for taking alternative beliefs seriously. Epistemic empathy and deep unbiased search for alternatives rise and fall together.

Predictions of existing psychological research literature. Existing research literature on AOT in adolescents predicts that there will be considerable variation and that it will predict grades and performance on cognitive tasks. Several related research programs support conflicting predictions about how widespread AOT is likely to be among adolescents.

The research literature on epistemological development suggests that sufficient epistemological sophistication to motivate AOT may be atypical among young adolescents. Based on series of interviews, most theories of naïve epistemology posit a gradual development of increasing sophistication, with many individuals remaining stuck in epistemological absolutism all their lives. Kuhn (2001) has argued that the final stage of epistemic development in her model—an evaluativist epistemic stance—is crucial to motivate actively open-minded thinking. For absolutists, only one view is right; for multiplists, all views are equally right. On either stance, learning about alternative views cannot add much to one's own. For an evaluativist, however, views have varying degrees of accuracy and thus much can be learned, even from partly wrong views. An evaluativist epistemic stance has not been found common among early

adolescents. Indeed, it was rare even in Perry's (1970) Harvard undergraduates. Although Kuhn set her bar for epistemic sophistication somewhat lower than Perry's, she, too, found only a minority of adults who ever achieved the evaluativist stance (Kuhn, 1991). Kuhn actually found an evaluativist stance slightly more often among those in their teens and twenties (13%-30%) than those in their 40s and 60s (8%-18%), although this difference is not statistically significant.

On the other hand, the evidence reviewed in Chapter 2 may indicate that adolescents are capable of much more sophisticated epistemological thinking than previous assessments showed. If this is true, then they may be just as inclined to actively open-minded thinking as adults.

The neo-Eriksonian research program on adolescent identity crisis hints that adolescents may be particularly good at actively open-minded thinking. Erikson (1968) described adolescence as a phase when many begin to question the beliefs and values of their parents, but have yet to fully commit to their own set of beliefs and values. This stage of openness and exploration may be prime for actively open-minded thinking; they have little to lose by being open to alternatives. Hence, the neo-Eriksonian adolescent identity literature could suggest that insofar as adolescence is the age *par excellence* for identity moratorium, many adolescents may exhibit quite high levels of AOT, although those in identity diffusion (confusion) or identity foreclosure (sticking unquestioningly to parents' views) may not (Grotevant & Von Korff, 2011; Marcia, 1966; Schwartz, Zamboanga, Luyckx, Meca, & Ritchie, 2013).

Identity foreclosure is associated with authoritarianism and intolerance of ambiguity (Berzonsky & Neimeyer, 1988; Schenkel & Marcia, 1972). Adolescents in either foreclosure or diffusion tend to have great difficulty dealing with conflicting information (Slugoski, Marcia, & Koopman, 1984). Moratorium and identity achievement, on the other hand, are associated with higher integrative complexity in social reasoning (Slugoski, Marcia, & Koopman, 1984;

Berzonsky & Neimeyer, 1988). Adolescents in moratorium are likely to be more open-minded because they do not think they have found answers to the big questions, but they are actively looking. Meanwhile, adolescents in identity diffusion are avoiding unresolved questions; they tend to be less information-oriented (Berzonsky, Cieciuch, Duriez, & Soenens, 2011) and more anxious (Schwarz, et al., 2011), which is likely to lead to less open-minded thinking.

To summarize, the existing literature makes no clear prediction for the extent to which adolescents value and engage in actively open-minded thinking. On the one hand, AOT theoretically requires a higher degree of cognitive and epistemological sophistication than is common among adolescents. On the other hand, adolescence is a phase of major transition without many commitments, which makes doubting beliefs and engaging in AOT less costly and perhaps even more valuable. Moreover, in Chapter 2, I offered evidence that adolescents have the capacity for more epistemological sophistication than is usually assumed. In light of these latter considerations, I predict that adolescents do value and engage in AOT to a considerable extent.

Contribution of present approach. Existing research on AOT does not reveal much about how thinking styles actually manifest in adolescence. In the past, AOT has been almost exclusively measured with a self-report rating scale, generally in conjunction with cognitive task measures of bias and cognitive ability for validation. Rating scales can efficiently assess a broad range of questions. However, their use in measuring intellectual personality traits requires that respondents possess considerable self-awareness of their own cognitive habits. There are inevitable demand artifacts and social desirability effects (see Duckworth & Yeager, 2015). All of these limitations are likely to be exacerbated among young adolescents.

Finally, subjects are likely to rate themselves using different comparison groups, rendering intergroup comparisons suspect (Heine, Lehman, Peng, & Greenholtz, 2002; Schwarz,

1999). Even in cases where the questions focus on norms and preferences, as with most AOT questions, many respondents are likely to exaggerate their commitment to habits that they know are likely to lead to better thinking. It is much easier to say "I listen to people I disagree with" than to actually listen to them. Such measures conflate valuing AOT with habits of engaging in it. Insofar as we are interested in both habits and values, they nonetheless remain valuable.

In the present study, we seek to explore the ways in which adolescents engage in actively open-minded thinking, using interviews and multiple-choice questions in addition to self-report rating scales. The addition of interviews and other non-self-report measures to traditional self-report rating scales can offer further insight into how people value and perform AOT. Interviews, particularly those in which participants are invited to think through problems to demonstrate their cognitive style, offer an opportunity to see people in the actual throes of thinking. It is thus possible to elicit demonstrations of capacities, instead of mere claims to possess desirable habits.

With this in mind, I developed a three-part semi-structured interview protocol around a set of problems and thought experiments. One was described in Chapter 2; the other two are the focus of the present work. These questions gave participants an opportunity to engage in both deep search for possibilities and evidence and epistemic empathy, across major differences of background and belief. The goal was to see what these capacities look like in action among contemporary American adolescents. Patterns emerging from the interviews could then be compared to patterns in the battery of quantitative data collected from a larger sample.

This battery, administered to a large, diverse group of eighth graders, included three types of self-report multiple choice items: (a.) several self-report rating scales in the classic form, including an 8-item Adolescent Actively Open-Minded Thinking Scale, (b.) a questionnaire gauging the extent to which understanding and truth were priorities, and (c.) a measure of

preferred ways of responding to disagreement. To remedy the limitations of self-report, we also collected multiple teacher ratings of every student and conducted extensive interviews with a diverse subset. The diversity of the sample was important for two reasons; to allow better generalization to the American population and to permit exploration of the effects of demographics on AOT. The study was thus designed to triangulate qualitative and quantitative data on the nature of actively open-minded thinking among American adolescents.

Part I: AOT Surveys

Study A. Broad Survey of Ninth Graders

Methods.

Participants. We recruited the entire eighth grade class from eight different schools in Pennsylvania, California, Texas, and Idaho, yielding a total of 1551 participants. The schools included two large, diverse public middle schools on the edge of a major city (constituting, respectively, 32% and 29% of our sample); four charter schools in three cities across the country (35% of our sample); and two small, high-end private schools (6% of our sample). Across the eight schools, 49% were female, 47% African-American, 25% Caucasian, 15% Hispanic, 11% Asian, and 2% Multiracial/Other. English language learners constituted 14%. Ages ranged from 11 to 16, but most were 13 or 14 years old (Mage=14 yrs 4 months, SD=5.7 months) at the time of the initial survey. Students qualifying for free or reduced lunch constituted 60%.

Because of time constraints, not all measures were administered to all participants. The Adolescent AOT Scale (AAOT) and Teacher Reported AOT were administered to 1551 participants and their teachers at all eight schools. Two exploratory but non-rating scale multiple choice measures were administered to a subset of 440, randomly assigned across students at all eight schools. Survey measures reported were administered near the end of the spring semester.

Procedure. All the students in one cohort at eight schools were asked to complete in-class computer questionnaires, twice in eighth grade (Fall Semester=T1, Spring=T2) and twice in ninth grade (Fall=T3, Spring=T4). Surveys were administered in classrooms by their regular teachers. An opt-out consent form allowed us to include circa 96% of students at the eight study schools. The survey was administered about a month before the end of each semester. Non-rating scale measures described below were included only twice, in the springs of 8th and 9th grade (T2 and T4). Interviews took place in the late fall of both years, in four of the schools.

Materials.

Adolescent AOT Scale (AAOT). We developed a novel rating scale to measure standards of actively open-minded thinking, the Adolescent Actively Open-Minded Thinking Scale (AAOT), based on an 8-item actively open-minded thinking developed for use with adults (Baron, Scott, Fincher, & Metz, 2014). The novel scale was piloted one-on-one and in small groups with students ages 12-13 to ensure adolescents would interpret the items as intended, and adjusted accordingly. Items focused on beliefs, preferences and norms of thinking rather than asking students to self-report their cognitive habits (which would require substantial self-awareness as well as honesty). For instance, one item, gauging beliefs about the utility of AOT, stated, "I can learn things by listening to people I disagree with." Others referred to cognitive norms, e.g. "People should pay attention to new possibilities," and "Changing your mind is a sign of weakness" (reverse scored). The response scale ranged from 1 = *strongly disagree* to 6 = *strongly agree*. (See Appendix E for full scale.)

Teacher Report AOT. The students' academic teachers at each school were asked to fill out a survey rating each of their students on five character strengths including AOT, based on four statements describing each character strength. The AOT item read, "paid attention to new possibilities; listened to people who disagreed with them; when others disagreed with them, tried

to understand why; and was willing to change their mind, given a good enough argument." Response scale ranged from 1 = "Not at all like this student," to 5 = "Very like this student." Ratings from up to three teachers were averaged for an overall AOT teacher rating.

Peer Nomination AOT. Every student was asked to nominate one member of their class or homeroom (depending on the size of the school) who was most likely to perform an action indicating each character strength. The prompt for nomination of a high AOT peer read, "Someone who will talk reasonably with other people even when they disagree, and try to understand why they think what they do." Because the peer nomination process was expected to produce a highly skewed distribution—most were not nominated at all, while some students were nominated by many—the number of nominations per student was log-transformed for the final measure of peer-nominated AOT.

*Argument Thoughts***.** The Argument Thoughts task (AT) is a measurement of openness to alternatives in the context of disagreement, consisting of three hypothetical argument scenarios. In the first, participants are asked to imagine having an argument with a friend. In the second, they are to imagine themselves observing an argument between two others in which they are inclined to take a side. In the third, they are asked to imagine themselves excited about an idea they have for a school project, only to meet with the skepticism of their assigned project partner. Below each scenario are listed 6-8 possible thoughts likely to occur to someone in that situation. The participants' task is simply to select the two or three of these thoughts most likely to occur to them.

The "thoughts" were generated based on the responses of 92 adults on MTurk each asked for three thoughts that often occur to them during arguments. Many of the thoughts generated in this way indicated high open-mindedness (e.g. "Try to empathize," "What is that other person

thinking?") or, more commonly, a lack of open-mindedness (e.g. "You're wrong," "I know I'm right," "I need to win"). Typical thoughts were written for each scenario such that half indicated high open-mindedness (e.g. "I want to understand what Arden means," "I wonder why they think that?") and half indicated low open-mindedness (e.g. "I'll prove you wrong!", "This is stupid; it's obvious who's right.").

The total Argument Thoughts score for each participant was calculated by subtracting the number of low-AOT thoughts selected from the number of high-AOT thoughts selected, across the three scenarios. Scores could thus range, in principle, from -8 (selected all low AOT thoughts and zero high AOT thoughts) to 8 (selected all high AOT thoughts and zero low AOT thoughts).

AOT Priorities. Because we were interested in how much the participants prioritized understanding and truth, we asked nine would-you-rather questions. Participants saw each question in a forced-choice paradigm, paralleling the form of the common game Would-You-Rather, where players must choose between options which are either both desirable or both undesirable.

Four *love of understanding* items pitted a love of truth and understanding against other desires like social status and comfort; three *openness to disagreement* items gauged preference for being challenged and learning something new versus having previous beliefs confirmed; and two *love of novelty* items gauged novelty-seeking. For example, we asked participants whether they would prefer to have a best friend who "Often disagrees with me" or "Agrees with me about almost everything." In this case, participants who expressed a preference for friends who disagree are expressing a preference for having their own ideas challenged, an indication of actively open-minded thinking.

AOT Norms. Many researchers have argued that a key technique for fostering habits of active, careful reasoning is engaging students in a classroom community with social norms of active, careful, open-minded reasoning (Brown, 1997; Cazden, 2004; Goos, 2004; Tishman, Jay, & Perkins, 1993). In order to pursue this hypothesis, in the spring of eighth grade students were given a 14-item AOT Norms scale, including seven items gauging perceived *injunctive norms,* that is, what behaviors and speech patterns peers approve or disapprove; and seven items gauging perceived *descriptive norms,* that is, how peers tend to actually behave and speak. The items in each subscale were matched; e.g. one pair of matched items read, "My classmates think it is bad to say that you are confused," (reverse-scored, injunctive norm), and "My classmates say when they are confused" (descriptive norm).

Scales for convergent validity. A large number of other rating scales were included in the large survey, mainly short versions of previously established scales.

Key for our current questions were two facets taken from Davis' (1980) sympathy scale: a 3-item *Perspective-Taking* scale, which includes items like, "When I am upset at someone, I usually try to put myself in their shoes for a while"; and a 5-item *Empathic Concern* scale, which includes items like, "Other people's suffering bothers me a lot." A single *Life Satisfaction* item was also included, as an outcome variable of interest. All rating scales other than the AAOT scale were on a five-point scale from *Strongly Disagree* to *Strongly Agree*, with no reverse-coded items. GPA was collected from schools and normalized by school as a second outcome measure.

Survey Results and Discussion

Descriptive Analyses.

AAOT Scale. The AAOT at first administration exhibited a Cronbach's alpha of .72 (N=1356) and a six-month test-retest validity of .59 (N=1306). It averaged 4.64 (SD=.69) with a range of 1-6, indicating that participants inclined considerably towards the actively open-minded

end, probably producing some ceiling effects. Over the 18-month span of the study, mean AAOT increased very slightly from 4.64 to 4.73 (*SD*=.69) (*t*=3.73 (922), *p*<.001).

Teacher AOT. Teacher-rated AOT fell along a normal curve, ranging from 1 to 5 with an initial mean of 3.28 (*SD*=.86), rising slightly across 18 months to 3.64 (*SD*=.87). Like the self-report AAOT, the teachers rated 60% of students above the midpoint, with modes at 3 (the midpoint) and 4, offering mild support for their students' high self-ratings. Furthermore, teacher-rated AOT correlated with self-rated AOT at *r*=.293 (*p*<.001). Teacher-rated AOT also increased slightly over the 18-month span, from 3.28 to 3.64 (*SD*=.87). This increase was significant (*t*=11.85 (1089), *p*<.001). Although teachers in eighth and ninth grade were different, test-retest correlation across the 18 month period was *r*=.47 (*p*<.001).

Peer-Nominated AOT. Peer AOT ranged from 0 nominations (56% of students in 8th grade, 53% in 9th) up to an outlier 18 nominations (the next highest was 9). Six-month test-retest correlation was *r*=.267 (*p*<.001), while 18-month test-retest correlation was *r*=.279 (*p*<.001), indicating some temporal stability. Peer-nominations probably suffer from some of the same halo effects as teacher-rated AOT, although this is harder to demonstrate because we did not permit students to choose the same peer as exemplary of multiple character traits. The predictive power of peer-nominated AOT is much weaker than teacher AOT. Presumably it is highly confounded with extraversion and popularity. Shy students could be very high in AOT but never nominated.

Argument Thoughts. Recall that Argument Thoughts scores could range from -8, for those who selected all low AOT thoughts, to 8, for those who selected all high AOT thoughts. Scores on this task among the 383 eighth graders who answered ran the full range, with a mean of 2.7 (*SD*=3.78), and mode and median both 4. Less than 1% of participants (four individuals) gave only low AOT answers (score of -8), while a full 14% chose only high AOT answers (score of 8).

Responses were similar in ninth grade, where median was 2, modes were 2 and 4, 61% chose predominantly high-AOT thoughts, and the mean was 1.8 (*SD*=3.17). The students thus indicated a fairly strong overall preference for applying AOT in disagreements. (See Table 8 for frequencies of each item.)

Does this indicate that adolescents actually do use sophisticated actively open-minded thinking when arguments arise? When one imagines a group of teenagers arguing, "I want to understand what you mean" is not a phrase one expects to hear. Yet 55% of eighth graders said they were likely to think just that. Because of time constraints, we gave participants a set of possible thoughts, rather than eliciting them in an open-ended paradigm. The form of the question was thus somewhat leading; it required only recognition of good thoughts, not their production. There are probably many students capable of recognizing high-AOT thoughts and remarks as wise who would complacently claim them as thoughts they would "be likely to think," but to whom such thoughts would never occur in the heat of a real argument. Despite our assurances of anonymity, doubtless some students were also influenced by what they thought their teachers would want them to say.

So if we cannot use this measure to conclude that eighth graders are consistently open-minded and thoughtful in actual arguments, what does it tell us? It tells us that they believe AOT is 'better' in arguments than dogmatism—perhaps because it is more productive or less harmful of relationships. At a minimum, however, it appears that on reflection, adolescents prefer AOT and prefer to think of themselves as expressing high AOT.

Table 8

Frequencies of Argument Thoughts

	T2 (*N*=447)	**T4 (*N*=953)**
Argument Scenario 1.		
1.2 I wonder why they think that?	48%	49%
1.3 Maybe we mean different things by that word.	34%	33%
1.4 Is there something they know that I don't?	40%	41%
1.5 What you are saying makes no sense.	48%	47%
1.6 Is there something we can agree about here?	35%	30%
1.7 I'll prove you wrong!	34%	39%
1.8 I know I'm right.	36%	33%
2.1 If I say something, I win because it'll be two against one.	22%	25%
2.2 Arden is being so stupid.	13%	17%
2.3 I want to understand what Arden means.	55%	52%
2.4 This is silly; it's obvious who's right.	38%	38%
2.5 Is there any way Arden could be right?	43%	44%
2.6 They're not listening to each other; how can I get them to actually listen?	73%	66.%
2.7 I want to make Arden look silly.	9%	10%
2.8 I'm not sure I understand the argument.	48%	45%
3.1 Just wait, I'll prove you wrong.	16%	16%
3.2 You just don't understand what I'm saying.	12%	20%
3.3 Why wouldn't it work?	59%	58%
3.4 Of course it will work! Just do what I tell you.	11%	13%
3.5 Will you help me figure out how to MAKE it work?	40%	33%
3.6 Wait, let me explain why I think it WILL work. Then you can tell me if you still disagree.	63%	57%

AOT Priorities. Items were scored 1 for participants who gave the high-AOT answer, otherwise 0, and summed for a total. Mean number of answers indicated love of truth/understanding was 6.2 out of 9 (*SD*=1.8, ranging from 0 to 9). The nine items did not hang together very tightly, yielding a Cronbach's alpha of .40, in part because the items were deliberately constructed to get at different aspects of AOT. Because of this, the descriptives for each item are given in Table 9. Participants responded most strongly to the items on love of understanding: 81% said they would rather be right but disbelieved than wrong but believed, 77% would rather understand an arithmetic problem than get it right, and 52% preferred understanding the material to getting a good grade.

For the five items repeated in spring of ninth grade, none were significantly different except for "Would you rather have a best friend who says when they think you're wrong, or always backs you up?" Choice of the high-AOT answer, "I would rather have a friend who always says when they think I'm wrong," increased from 52% to 64% (t=5.55 (259), p<.001) across twelve months.

These results indicate substantial variation in prioritization of understanding over competing interests. Middle school is a stressful period; many students feel pressure from their parents to get good grades and sometimes conflicting social pressure from their peers, and few have yet developed effective coping strategies (Wenz-Gross, Siperstein, Untch, & Widaman, 1997). It is possible that some of those who prefer agreeable friends do so because they desire emotional support over intellectual stimulation from their friends. These other priorities are therefore not surprising. Nonetheless, a considerable number chose understanding and challenge over these other strong desires. On the whole, the AOT Priorities measure indicates that a love of understanding is widespread, although sometimes overpowered by other goals.

Table 9

Frequencies of AOT Priorities

Set	Item	% choosing AOT answer		
	"Would you rather….."	8th Grade	9th Grade	Test-Retest *r*
U	Be right vs. be believed	81%	--	
	Understand math vs. get it correct	77%	78%	.19*
	Understand material vs. good grade	52%	47%	.31**
	Learn about friend vs. win argument	68%	62%	.34**
C	Friend who says when wrong vs. backs me up	52%	64%	.27**
	Friend who disagrees vs. friend who agrees	39%	50%	.31**
	Fix a mistake vs. confirmation of being right	45%	--	
N	Book about Mongolia vs. US	53%	--	
	Book about other religion vs. family's religion	53%	--	

Notes: ***p*<.001; **p*<.01. U indicates item on love of understanding; C indicates item on love of intellectual challenge; N indicates love of novelty. *N*=441.

AOT Norms. The items on the two AOT Norms Scales, Descriptive and Injunctive, fell together in a factor analysis and fell into the same part of the nomological net—that is, anything one predicted, the other predicted to about the same degree. For this reason, I combined them into one AOT Norms scale. Together, the two scales showed acceptable internal reliability, with a Cronbach's alpha of .78. The AOT Norms composite scale was normally distributed, ranging from 1.44 to 4.69, with a mean of 3.15 (*SD*=.47).

Correlations. Partial correlations were computed controlling for school using dummy variables (see Table 10). However, results shown are robust to removing all controls, or adding controls for ethnicity, gender, and free/reduced lunch as well as school (although in the latter case

correlations are somewhat attenuated). The AAOT showed convergent validity with the teacher-reported AOT (r=.27, p<.001, N=1324) and the two behavioral tasks, Argument Thoughts (r=.44, p<.001, N=341) and Priorities (r=.33, p<.001, N=429), indicating operational validity of all three measures (see Table 11). Consistent with the AOT construct, the AAOT scale also predicted the empathic concern scale (r = .41, p<.001, N=1185) and the perspective-taking scale (r=.42, p<.001, N=1302).

Table 10

AOT Partial Correlations (Spring 8th Grade)

	1	2	3	4	5	6	7	8
1. AAOT	--							
2. Teacher-rated AOT	.19**	--						
3. Log Peer-nom AOT	.11*	.23**	--					
4. Argument Thoughts	.44**	.27**	.22**	--				
5. Priorities	.32**	.25**	.03	.43**	--			
6. AOT Norms	.21**	.13*	.06	.12*	.17**	--		
7. Perspective-Taking	.37**	.17**	.06	.21**	.17**	.11*	--	
8. Empathy Scale	.38**	.22**	.12*	.19**	.14**	.07	.43**	--
9. GPA	.23**	.59**	.15**	.17*	.10	.08	.06	.17**
10. Life Satisfaction	.15**	.04	-.03	-.01	.02	.21**	-.02	.02

* indicates p<.05; **indicates p<.01, both 2-tailed. 3. Log Peer-nom AOT was the log-transformed number of times each student was nominated as an AOT exemplar by peers, to normalize. Partial correlations shown controlling for school. All measures above are from the second survey wave in Spring 2015, which was the first time measures 3-6 were given. N > 429 for each; actual sample size varies depending on missing data.

Supporting the value of AOT for some life outcomes, the AAOT scale predicted Life Satisfaction (r = .15, p<.001) and GPA (r = .08, p<.05). Argument Thoughts also supported the construct's predictive power for GPA (r = .17, p<.05). The much stronger correlation between GPA and teacher-rated AOT (r = .59, p<.011) should be taken with a grain of salt, since halo effects for teachers are well established (e.g. Foster & Ysseldyke, 1976). It is not surprising that

teacher ratings of AOT correspond so closely with grades, since both teacher-rated AOT and GPA require the same teachers to rate the students' intellectual qualities. Nonetheless, this finding does suggest that teachers consider actively open-minded thinking a strength in their students, which is encouraging.

The two tasks that corresponded most substantially with the AAOT, Argument Thoughts and Priorities, were not classic self-report rating scales, but, like the AAOT rating scale, did involve self-report of preferences and inclinations. This similarity in measure type is probably why these two measures showed the greatest correspondence with the AAOT. Indeed, Duckworth and Kern's (2011) meta-analysis of self-control measures found that 53% of the variance in effect sizes was accounted for by measurement type, with the greatest consistency within self-report scales.

The cognitive style of actively open-minded thinking is not frequently contemplated by your average adolescent. Thus, it was not possible to ask about attitudes and behaviors quite as concrete as one could for, e.g., a study of sexual attitudes and behaviors. This unavoidable lack of concreteness and specificity almost certainly attenuated all correlations in the present study.

Gender. Girls in our sample exhibited significantly higher levels of actively open-minded thinking, not only in self-reported AOT (t=5.45 (1291), p<.001) but also in Teacher Reported AOT (t=7.03 (1392), p<.001), Argument Thoughts (t=4.82 (442), p<.001), and AOT Priorities (t=2.31 (424), p<.05). The effect size of this gender difference in AOT is comparable to the gender difference in GPA, where girls also have a significant advantage, or the difference in life satisfaction, where girls are at a disadvantage (see Table 11). Girls also reported higher empathy and perspective-taking. These gender differences appeared, at least directionally, within every ethnicity and in every school except for one private school, which had a small sample size of 50.

Moreover, the gender differences did not change significantly over the 18-month period, although AOT rose slightly for both genders on every measure. The only measures related to AOT which did not show a significant gender difference were AOT norms and peer-nominated AOT. This is not likely to be a rating scale artifact of girls agreeing more than boys, because, out of several dozen self-report rating scales, AOT was the only one which included reverse-coded items, and it was one of very few which showed any gender difference at all. The effect size of the gender on AOT ranged from *d*=.2 to *d*=.45, depending on the measure. This was comparable to the advantage of girls in GPA and disadvantage in life satisfaction, both of which should be particularly pronounced at this age.

Table 113

AOT Gender Differences

Spring 8th Grade	***t (df)***	***Cohen's d***
Self-Report AOT	5.45*** (1291)	0.30
Teacher-rated AOT	7.03*** (1392)	0.35
Peer-nom AOT	0.5 (1492)	0.03
Argument Thoughts	4.82*** (442)	0.45
AOT Priorities	2.31* (424)	0.22
AOT Norms	1.04 (430)	0.11
Life Satisfaction	-6.45*** (1304)	-0.35
GPA	7.59*** (1353)	0.41
Empathy	7.23*** (1299)	0.40
Perspective-Taking	2.36* (1298)	0.13

Note: ***p<.001; **p<.01; *p<.05

The gender difference might be a result of differences in maturity, which are at their maximum in middle school. Puberty strikes girls between 8 and 14 years, boys not until 10 to 16. Puberty is a time of considerable cognitive change (Sisk & Zehr, 2005). However, the absence of

any correspondence between age and AOT (r=.003, N=1337, *ns*) counts against this interpretation.

Alternatively, it may be a result of differences in social intelligence, interest in the thoughts and feelings of others, and/or epistemic humility. There is considerable evidence that girls tend to be more empathic and socially intelligent than boys, a difference which persists through adulthood (Eisenberg & Lennon, 1983). Boys tend to be more confident than girls (Lenney, 1977; Lirgg, 1991), which might manifest in an unwillingness to listen to disagreement, perhaps from some norm of machismo which is enforced more strongly among boys.

However, despite the robust gender difference across measures of AOT, the gender difference in AOT Norms was not significant (t=-1.04 (430), *ns*). This does not necessarily rule out the possibility that some boys are responding to norms of machismo, since the AOT Norms items did not ask students to distinguish between female norms and male norms. Nonetheless, it does suggest that the gender difference in AOT is not merely a result of different norms or perceptions of norms for girls vs. boys.

Ethnicity. The possibility that AOT would appear differently within different ethnicities was tested using linear regressions including school, gender, English Learner status, and ethnicity as dummy variables. This analysis found no statistically significant effect of any ethnicity on self-reported AOT at any of the four time points, except for a marginally higher average from Asian-Americans at T1 and T2 (at both T1 and T2, β=.08, $t\approx 2$, p=.04). There was also no difference across ethnicities at either grade level for scores on AOT Priorities (all ts ≤ 1, ps>.05), AOT Norms (all ts<2, ps>.05) or peer-nominations for AOT (all ts<2, ps>.05, again with a close exception for Asian-Americans at T2, who were selected marginally more often as high-AOT, t=1.97, p=.049). Argument Thoughts also showed no ethnic differences except that Asian-

American students chose more high-AOT thoughts at T4 (t=2.85 (363), p=.005), though not T2. Thus, it appeared that AOT was consistently prized across all ethnicities.

However, teacher-rated AOT did differ by ethnicity, with teachers rating Asians higher (β=.20, t=5.80, p<.001) and African-Americans lower (β=-.21, t=-4.98, p<.001). English language learners were also rated slightly lower than their classmates (β=-.07, t=-2.45, p=.02).

This discrepancy between teacher-reported AOT and the other measures could mean the teachers were stereotyping based on race. It could also arise more legitimately from the teachers picking up on different aspects of enacted AOT than self-report or peer-nomination, for instance how AOT is demonstrated in schoolwork rather than conversation.

It is somewhat encouraging that peer-nominations were apparently blind to ethnicity, suggesting that insofar as ethnic stereotypes may have played a role in these judgments, it was limited to the older generation. It is possible that the apparent lack of ethnic bias in the students arose from in-group bias, such that students primarily chose other members of their own ethnicity, on account of ethnically segregated friend-groups. However, the large majority of the schools were extremely diverse and in interviews many spoke of friends of other ethnicities, so this is unlikely to be the whole story.

It is striking that the valuing of AOT seemed to be constant across all ethnic groups. This is unlikely to be due to insensitivity of the measures, since they did pick up consistent gender differences. Ethnic differences in success in school are commonly found, and replicated in this sample for GPA. In a regression with dummy variables for school, gender, and ethnicity on GPA at T1, African American students tended to have lower GPAs (β=-.24, t=5.68, p<.001) and Asian American students tended to have higher GPAs (β=-.23, t=6.75, p<.001). The value of actively open-minded thinking, however, seemed to be shared out equally.

Study B: Narrow Survey of Sixth Graders

To corroborate the finding that middle schoolers across demographics put high stock in AOT, the AOT self-report measure was also given to 177 sixth graders (48% female) in three low-income public middle schools in a large East Coast city, along with the same self-report measures of perspective-taking and empathy used in Study 1a. In support of the construct validity of the AOT self-report measure, among the sixth graders, AOT again predicted the theoretically related measures when controlling for gender and school, including perspective-taking (r=.42, p<.001), empathy (r=.60, p<.001), and love of reading (r=.33, p<.001). AOT also predicted GPA z-scored by school (r=.28, p<.001).

Most strikingly, the distribution of self-reported AOT in this very low privilege population two years younger was almost identical to that in the 8th graders across schools. The mean AOT score for the inner city sixth graders was 4.58 (SD=.75), which is not significantly lower than that of the 8th graders (M=4.64, SD=.69).

Taken as a whole, the survey data from sixth, eighth, and ninth grade offer several converging pieces of evidence that contemporary American adolescents across differences of age, school, SES, and ethnicity care about AOT and want to engage in it. Many of them (a.) endorsed norms of actively open-minded thinking, including the importance of listening to those who disagree and seeking out new information; (b.) expressed a preference for friends who sometimes disagreed, indicating a willingness to put pressure on their beliefs; and (c.) said they try to understand an opponent's perspective in the context of an argument. Even if they are not reporting their behavior honestly, this evidence indicates at a mimimum that valuing AOT is widespread.

The 96 interviewees expand upon this evidence, demonstrating what the adolescents' AOT looks like in action. As we will now see, these interviews indicated that many of our participants did not only wish to be actively open-minded, but genuinely *were*.

Part II: AOT Interviews

Methods.

Population. The interviews were administered in the fall semester of eighth grade to a subset of 96 participants distributed across the two public schools, three of the charters, and one of the private schools (M_{age}=13 yrs 10 months, SD=5.6 months). Interviewees were chosen by school administrators to represent top, middle, and bottom achievers at each school. The resulting 96 interviewees consisted of 45 African-American, 24 Caucasian, 15 Asian-American, nine Hispanic, and three Multi-racial participants, with 51 students on free or reduced lunch, 53 females and 43 males, and 12 English language learners who had attained sufficient fluency for the interview. Nine attended the private school, 39 one of two public schools, and 48 one of four charter schools. Eight attended a charter school which was not included in the survey study.

Interview procedure. Each participant was interviewed one-on-one by a trained interviewer in a room in their school during the school day. The AOT portion of the interview came after 20-40 minutes of questions about the interests, habits, and preferences of the participant, which usually made the participant engaged and comfortable with the interviewer. Discussion of the protocol and results of the interviews are divided into two key sections: (I.) Problem Solving, which sought to elicit manifestations of participants' capacity for engaging in deep and open search and consideration of alternatives, and (II.) Epistemic Empathy, which used two scenarios designed to challenge participants' capacities for epistemic empathy. Each section below describes the interview protocol, the coding categories, and emergent patterns in adolescent cognition.

Interview analysis. Each interview was coded according to a rubric of ideas or remarks which were either mentioned (coded 1) or not mentioned (coded 0). The first 20 interviews were coded by two coders; Cohen's kappas ranging from .61 to 1 were achieved, depending on the code, with a mean of .75 (Cohen, 1960). Thus all codes were within the range of "substantial agreement" identified by Landis and Koch (1977) and the mean at the cusp between "good" and "excellent" agreement (Fleiss, 1981). The remaining 76 interviews were coded by one coder. All coding was blind to school, demographics, and quantitative scores.

Section I: Depth of Search.

Protocol. This section was designed primarily to provide an opportunity to observe the depth, breadth, and openness of each participant's search for possibilities and evidence. We sought to elicit particular ideas to reflect high or low AOT (see Table 12). To this end, students were asked to (a.) develop a solution to a problem, (b.) describe how they would search for information and alternatives, and (c.) respond to hypothetical disagreement from a peer.

Specifically, interviewers invited students to imagine taking responsibility for creating an eighth grade science curriculum of their choice. This topic was selected because of its familiarity to all students and intrinsic interest, since they could easily imagine themselves being affected by it. To interest them further, the interviewer began by asking about their favorite topic in science:

What part of science do you find most interesting?

Next, participants were asked to describe how they would *figure out* what and how science should be taught, to gauge the depth and sophistication of their search for information:

> *Suppose you are an advisor to the school principal at* [student's school], *and* ***you*** *get to decide what is going to be taught in eighth grade science, and how it will be taught. How would you go about figuring out what eighth graders should learn about* [favorite topic]*? Are there any other ways you would figure out what the best things to learn are?*

Next, they were asked what forms the curriculum might take, as a measure of depth of search for possibilities.

> *What do you think students might learn about [favorite topic], and how would they learn it?*

After students were fairly invested in their idea by all this elaboration, they were asked to imagine a fellow student coming along who didn't like their idea. After giving their reaction to this dissenting classmate, participants were asked why that classmate might not like the idea, in an attempt to gauge whether the participant thought disagreement could be legitimate:

> *If a student at your school said to you, "I think your ideas about how to teach science are completely wrong," what would you do or say? Why do you think they might have said that?*

Table 12

Coding Schema for AOT Section I: Depth & Breadth of Search

Code	*Examples*
Proposes testing out alternative ideas (high AOT)	[*I: What would help you decide whether to input* [the other student's idea*]?]* S: Maybe trying them out in a classroom environment and see how they work.
Asks for dissenter's ideas (high AOT)	"I would try to talk to them and see what their idea would be." "I would ask him what he thinks it should be like."
Willing to change own idea (high AOT)	"I would change it, and I would get feedback from students and see what they want – what they – what's wrong – what am I doing wrong."
Reluctance to change idea (low AOT)	"I would not take their opinion."
Offended by dissent (low AOT)	"That was kind of mean." "Cause maybe he doesn't like…me."

Results and Discussion. Some students demonstrated little information search and offered few possibilities, while others said they would ask many sources (students, teachers, textbooks, Google, etc.) and considered many possible modes of learning (lecture, games, experiments, films, books, etc.) (See Table 13 for summary of key results). Fully 52% (51 students) said they would consult other students, usually via a survey; 34% (34) said they'd ask an expert like a good teacher or educator. The least sophisticated information sources were their own past experience (17%, 16 students) (e.g. "Probably from previous years on what you learn and how you felt about some stuff") or personal preferences (15%, 14 students) (e.g. "I would like to [do] life science because that's my favorite").

Although the modal number of sources for information search was one (27%) or two (28%), the remaining third mentioned three or more, a fairly broad search for information.

Table 13

Common Themes in AOT Section I: Depth & Breadth of Search

Codes	***# Interviews***	***%***
Proposes testing out alternative ideas	13	14%
Asks for dissenter's ideas	57	59%
Willing to change own idea	42	44%
Reluctance to change idea	32	32%
Offended by dissent	18	19%

Notes: Not all percentages are out of the same total, since parts of some interviews were uncodeable due to interviewer error.

In response to a hypothetical student's dislike of their idea, 59% responded openly by asking for the dissenter's ideas with the hope of improving upon their own, indicating a high degree of openness. Forty-four percent voluntarily and explicitly said they were willing to change

their curricula if the dissenter had good ideas. Only 19% responded with any trace of resentment or offense for the dissent, e.g. "[He doesn't like my idea] 'cause maybe he doesn't like me." This 19% tended to be suspicious of the motives of the dissenter, suggesting that their offense was derived from past experience with smart-aleck dissension rather than any deeper unwillingness to consider alternatives.

On the whole, the Depth of Search section indicated a common though not universal willingness to engage in moderately deep search for information and possibilities, and a fairly common openness to disagreement and alternatives, although a large minority of interviewees were uninterested in the dissenter's ideas.

Section II: Epistemic Empathy.

Protocol. To probe the students' capacity for epistemic empathy, the second section asked them to engage in two thought experiments, Different Family and Women's Vote (see Table 14 for summary of codes).

*Interview Section IIa. Different Family***.** The Different Family section asked students to entertain the hypothetical that they had grown up in a completely different family in a different place. Interviewers then asked if any of their beliefs might be different in that other life:

> *The next question is about your beliefs and opinions generally. If you had grown up in a different family in a different town, might you believe some things you don't believe now, or would you believe all the same things?* (If yes) *Like what? Why/Why not? Might you not believe some things you do believe now, or would you believe all the same things?* (If yes) *Like what? Why/Why not?*

This question was designed to elicit participants' awareness of the contingency of their own beliefs on circumstances. An epistemically empathic individual would recognize that even her own beliefs were largely a result of the teachings and mores of family and community, and as such, could easily have been otherwise. This recognition might lead her to consider which aspects

of her own upbringing might have affected her beliefs without truth-tracking – that is, *causes* of belief which were not valid *justifications* for belief – which in turn would lead to a critical examination of her own beliefs.

Table 14

Coding Schema for AOT Section II: Epistemic Empathy

Code	*Examples*
Pluralist Values	"To get a second – if all men were able – they were just able to vote, then they would just have the opinion of men, but you need different people to put their opinion – like the women's opinion."
Talk of human rights (deontology)	[I: (*Following up on student's claim that women should have the vote)* Why is it better if women can vote?] S: "Because she's still human. She has her rights. It's America. You have your rights."
If family diff, beliefs could be diff	"I guess because families have different viewpoints. There's a lot of different opinions on how many things were created, I guess."
If family diff, religion could be diff	"A different family, I could be in a completely different religion. Maybe my beliefs – my mom's Catholic. I live with my mom, so I'm Catholic. I believe in Catholic. But if I was in a Jewish family, my beliefs change. I believe in the Jewish guy. I would worship different."
If family diff, a value could be diff	**"**Oh, I would most likely think different things, like I have very liberal … and I think if I grew up in a different part of the country, I might feel differently about things like that."

Nothing could be less dogmatic and more actively open-minded than a critical examination of one's own beliefs. Still more easily, in imagining how she herself might have believed otherwise, she would recognize her fundamental similarity to those who hold different beliefs, and the temptations to dogmatism and prejudice would fade. A more dogmatic individual must suppose his line to truth fairly reliable, and thus not dependent on arbitrary circumstances like where he happened to be born.

Section IIb. Women's Vote. Women's Vote, the second thought experiment, asked participants to consider the time before women's suffrage:

> *People used to believe that women should not be allowed to vote. Why do you think they believed this? Would it have been better for society then if women were allowed to vote? Why/Why not?*

Finally, they were asked to consider the perspective of a particular man of that era, husband and father to women:

> *Imagine a man with a wife, a son, and two daughters who lived in 1850, when women couldn't vote. He believed women shouldn't be allowed to vote. Do you think it was reasonable for him to believe that, or unreasonable? Why/Why not?*
> (If unreasonable) *Did he have any reasons for his belief that women shouldn't vote?*

Epistemic empathy involves recognizing the reasons another has for their belief, even a detested belief. Students were asked *if* such a man would have any reasons rather than *what* his reasons might be in order to avoid pushing them too much into inventing reasons from perceived experimenter demand.

Results and Discussion. Many students demonstrated considerable epistemic empathy in a variety of ways. (See Table 15 for a summary of common themes.)

Different Family. Of the 96 interviewees, 76% admitted they would have some different beliefs, and many were able to specify beliefs that would be different. The most frequent contingent beliefs they mentioned were religion (42 students; 45%) and values (25 students; 27%).

Recent research suggests that values are the most central aspect of identity, at least for adults (Strohminger and Nichols, 2014), although plausibly the values adolescents said came from their parents were not those at the center of their identity (e.g. value of education or gratitude). The belief that one's own religion is supreme has frequently motivated dogmatism and

bigotry throughout human history. Yet here are these 42 eighth graders frankly volunteering the fact that their own religion is contingent on their upbringing, and 25 saying some of their values are, too. This went in many directions; Christians spoke of the possibility of being atheists, Muslims and atheists of the possibility of being Christians. A creationist spoke of the possibility of accepting evolution, were her parents different, saying:

> A lot of people's opinions and things they think has to do with who they grew up around, who they're around. If I went to a school that was saying that we had—say my parents went to that school, too, and they grew up being told that humans came from monkeys, and they would tell me that, too, and the school I went to would tell me that, I think that would make me think that they did. I probably wouldn't disagree with it.

Many participants were able to imagine believing otherwise about the most important things. They already understood a key insight from one of the greatest skeptical essays of history, Michel de Montaigne's *Apology for Raymond Sebond.* In 1568, while the Catholic and Protestant wars raged, Montaigne wrote, "We receive our religion in our own way and through our own hands, and no differently from the ways in which other religions are received . . . [in other circumstances] another religion, other witnesses, similar promises and threats could in the same way imprint in us a contrary belief." Although this insight may have been shocking and controversial in Montaigne's time, it seems that many contemporary American eighth graders already possess this insight. Plausibly this is a result of the multiculturalism of American schools and communities. It must be far easier for a Jewish child to imagine what it is like to believe in Mohammed and Allah if one of her friends is Muslim.

What consequences this may have for religious dogmatism have yet to be seen. Nevertheless, it seems plausible that a child who can so easily imagine accepting a different religion would be unlikely to condemn the followers of that religion quite as fiercely as her ancestors. Multiculturalism may be transforming the nature of religion in ways we have yet to fathom. Even the Pope recently implied that it might be possible for atheists to get into heaven,

saying, “Sin, even for those who have no faith, exists when people disobey their conscience” (Day, 2013). Perhaps the increased exposure to other views that comes with multiculturalism increases the capacity for epistemic empathy.

Women’s Vote. Epistemic empathy was not limited to beliefs the students had encountered before. Almost no one now will say in public that women should not be allowed to vote. Yet 78% of participants were able to come up with some kind of reason a man in 1850 might have refused women the vote, despite thinking him wrong. Twenty-two percent said the man did not want women to vote because of his upbringing or culture, which they saw as excusing rather than justifying. For instance, Darren said, “That man was raised to know women aren't allowed to vote, so he went along with it.”

Table 15

Common Themes in AOT Section II: Epistemic Empathy

Codes	# Interviews	%
If family diff, beliefs could be diff	73	76%
If family diff, religion could be diff	42	44%
If family diff, a value could be diff	25	26%
Sexism ascribed to culture/upbringing	21	22%
Pluralist Values	29	30%
Talk of human rights (deontology)	14	18%
Other		
Spontaneous expression of AOT norms	9	9%

Notes: Not all percentages are out of the same total, since parts of some interviews were uncodeable due to interviewer error. “Diff” is an abbreviation for “different.”

They may have been primed for the influence of upbringing by the previous question about growing up with different family. Tom made this explicit:

> I think this ties back to the whole upbringing thing. If he was raised to believe that, and he doesn't put much thought into this 'cause he's got a busy life…If this guy's a pioneer guy, you can't really expect him to do a lot of thinking all the time. So people usually just did what they were taught back then, and they used it as a code to live their lives.

However, many participants were able to offer other reasons. Seven students (7%) pointed out that in 1850, women were far less educated than men, and thus likely to vote less intelligently. Darren, for instance, said, "I think they usually thought that because women did most of the work at home and they didn't have that much of an education."

Indeed, although 41% (39) of participants said explicitly that a historical man would be "unreasonable" to deny women the vote, most could describe plenty of reasons. 'Unreasonable' seemed to indicate disapproval of his position rather than an indictment of the man's reasoning. One interpretation is that those who called the man 'reasonable' were relativists; within his time and place, they suggested, it was appropriate to think women couldn't vote. In fact, several affirmed this, like Izzy, who said, "I don't know. It's just really a matter of opinion. In his opinion he didn't think women should vote. That was just what his opinion was. I think everyone should have an opinion." They seemed to be moral relativists, if not epistemic multiplists. Yet even in the moral case, most went on to say something indicating they were not true moral relativists. Izzy continued, "But I don't think that was fair."

Most participants in this study were not only able to entertain both thought experiments without any trouble, but demonstrated considerable epistemic empathy in both Different Family and Women's Vote.

Other patterns emerging from interviews.

Norms of Pluralism. The adolescents repeatedly expressed norms of pluralism, primarily in the context of the Women's Vote subsection. By *pluralism,* I mean the idea that multiple

perspectives are in some way legitimate, without necessarily claiming that all perspectives are equally legitimate (which latter would constitute what Kuhn calls the multiplist epistemic stance).

Eighty-three percent (79 students) said that it would have been better if women were allowed to vote. Those who said it might not have been better were generally concerned with the unpredictability of how a big change in the past might affect the present. This was put most trenchantly by Trevor, who said, "You don't mess with a time machine." This worry, however, was held by a minority.

Fully thirty percent (29 students) said it would be better if women could vote in light of the value of pluralism, saying something along the lines of "more opinions is better." These pluralist values were justified in two ways: of pluralists, 21% (6 students) justified pluralism on a *deontological* basis of rights—"every [one] has a right to their own opinion"—and 38% (11 students) justified it on the *consequentialist* basis that more opinions would lead to a better or more accurate outcome—e.g., "If we had mixed opinions it would be accurate."

Deontological Justifications for Pluralism. Expression of deontology-grounded pluralism seemed to be based in some background assumptions about human rights, grounded more in considerations of subjective beliefs and desires rather than in a hope for objective truth. These participants emphasized the personhood of each individual, giving everyone "a right to their own opinion." Hal, for instance, said, "Everybody is a person and deserves to do what everybody else do" [sic]. Some focused on the right to expression. Naeela said, "Everybody's opinion matters…There's a whole 'nother gender that needs their opinion to be voiced." Yara emphasized the right to pursue one's own ends as well as the right to expression, saying, "Because even if they don't agree on the same person that they think should win, she should still get her opinion because everybody get to tell their opinions and do what they wanna do."

Many spoke of the importance of getting multiple perspectives to make it more "fair" or "even," like Ashley, who said,

> Yeah, 'cause then you're getting more people to put what you actually think should happen. 'Cause if there's a presidential nominee that all men like, but not all women did, they could persuade the voters and make it more fair or even.

Some appealed to the value of diversity along with equality. Juan offered the most articulate deontological justification for pluralism:

> I think you always need an equal, fair thing, like if you always have one race vote on everything, one gender vote on everything, then it's not gonna be diverse. The people who are not what everyone's being voted on don't have a say. I feel like you can't really, you can't really vote on something with your heart unless you've experienced it, and if people haven't experienced what it's like to live in another person's shoes, then I don't think they can really push those folks to do something.

This family of deontological justifications for pluralism sometimes might be taken to imply the consequentialist justification, but not always. A few of these participants suggested that even though it would have been fairer, it would not necessarily have led to a better outcome if women had had the vote back in 1850. Anna was optimistic that her values of fairness and consequences would not have conflicted, but not entirely sure: "I think it would have been fairer to the people. I don't know if history would have gone better. Probably yes because there would have been more equal representation of people in the vote."

Two boys seemed to take the right to one's own opinion all the way to moral relativism. Darius said, "It was reasonable [for the man to be against women voting]. Because everybody has their different opinions. One person might believe that they should be able to vote, while another person might believe that they shouldn't." The simple fact of different opinions seemed to be self-justifying for him. Paul was more explicit, "Back then, that's something you believed that happened. Owning slaves back then wasn't wrong. It's just something you should have. That's

what they believed back then, but that all changed." The example Paul chose—slavery—is the very one that tends to make relativism-inclined undergraduates balk.

To summarize, pluralism was often justified for the sake of fair, equal, and diverse representation, goals which seemed to be valued for their own sake.

Consequentialist justifications for pluralism. Consequentialist justifications were based in the hope that more opinions were not only fairer and more diverse, but would for those very reasons help the community come to an objectively better or truer outcome. For example, Dan justified giving women the vote because "Usually you would also want to have a mixed representation of the whole country, this is probably more in the voting kind of thing. If we had mixed opinions it would be accurate." That is, he thought a more diverse and representative set of voters would lead to better, "more accurate" election results. Rick put it more simply: "It's more opinions to help you out with the decision."

Some of them seemed to be think that a wider variety of people voting would improve outcomes through the processes of reflective democracy, where a vote was a not just a vote but a voice with its own ideas and suggestions. Maria said,

> Every voice matters and that one vote can maybe make a big difference. Because if it was just all men, maybe they all agree, but I guess with the women it would be different. They would be outnumbered and they're different. Each vote gets a different thought about a situation, so they can more clearly look at it and everything.

Although a vote itself isn't enough to constitute "a different thought," Maria is right to think of getting a vote as symbolic of getting more of a voice in broader political discourse. She seems to have in mind something like J. S. Mill and Oliver Wendell Holmes's ideal of a marketplace of ideas, where freedom of speech for everyone gives the best ideas a chance to be heard and recognized. Tanya put the virtues of diversity in political discourse even more optimistically:

> Because looking at today, there are so many new ideas brought up I guess and so many new things being challenged and so many new things being thought about that weren't thought about when it was just men. And having men and women, there's double the intelligence and double the chance of a society succeeding or something like that. So it's just like doubling your chances. Why wouldn't you want that?

Tanya's appreciation of "so many new ideas" and "so many new things being challenged" is only possible for someone who strongly values AOT.

Indeed, a respect for pluralism goes hand in hand with the belief that AOT is the best way to achieve true beliefs and good results. This is especially true of a pluralism justified by consequentialist considerations. The more we think that something can be learned from a variety of perspectives, the more open we will be to perspectives that are not our own. The pluralistic norms in many corners of the contemporary multi-cultural United States may even be the reason adolescents put as much stock in AOT as they do.

Spontaneous Expression of AOT Norms. Nine participants (9%) actually offered spontaneous expressions of norms of actively open-minded thinking. This occurred at least once in each of the two sections; two participants expressed such ideas in two different sections. Wanda, describing her response to a dissenter in Section I, even described engaging in epistemic empathy:

> I would listen to their ideas, since it's always kind of wrong to always think that your ideas are right. Then I would think of how their idea is good, and how it's also bad...Well, I would try to be that person and think about that idea and pretend I like it...I would sort of agree with him, and I would try to find a middle point between their idea and my idea.

Note that this epistemic empathy does not lead her to abandon her original idea, or even to consider the dissenter's idea uncritically. Indeed, she says explicitly that she will "think of how their idea is good, and how it's also bad." Another student, Andrew, discoursed freely upon the virtues of actively open-minded thinking in conversations, saying:

> I think the problem with having debates – and I think in class, too, if we're having a debate in humanities – the problem with debates is that people get so wrapped up in their opinion that they kind of refuse to consider anybody else's because it becomes less let's talk about these ideas and more my opinion versus your opinion, me versus you. This is an argument. I don't – that's why I don't like debating stuff. I love discussing opinions, but to me, that's really different from debating.

Both Wanda and Andrew recognize the effort involved in actively open-minded thinking, but assert forcefully and lucidly their desire to do it anyway, even when their conversational partners may not be on the same page. Of the nine eighth graders who made such a statement, only Andrew was at the private school. These nine adolescents volunteered norms of actively open-minded thinking unbidden, when they were merely asked to talk about expert disagreement or being faced with dissent from their own ideas. Although nine is a small minority of the 96, these unsought declarations constitute further evidence that many eighth graders are eager to engage in actively open-minded thinking.

Conclusion

Ninety-six interviews and surveys of 1,551 adolescents from public, charter, and private schools across the country demonstrated unexpectedly widespread capacities for actively open-minded thinking, undergirded by considerable epistemic empathy and a strong value of open-mindedness. The large majority of these 13- and 14-year olds repeatedly expressed openness to disagreement, awareness of their own ignorance, and a desire for real understanding. Although there was certainly variability in the sample, the majority of participants showed high actively open-minded thinking across both quantitative and qualitative measures. They also demonstrated a widespread capacity for epistemic empathy, as demonstrated by their grasp of the contingency of their own beliefs and values even for the most strongly held.

It took months of focus-grouping to find questions which would goad some adolescents into expressing close-mindedness. Indeed, the overriding challenge in developing the measures was how to scrape the adolescents off the ceiling. Even in the final forms of the measures, the

distributions still skewed heavily left. Were this only the case for the self-report measures, this might be a mere artifact of response desirability. It would show that the participants knew open-mindedness is valued in our culture, but it would not show they actually had the capacity for it. However, the teacher ratings and interviews back up the self-reports. One after another, interviewees invited advice from someone who didn't like their ideas (59%), admitted the contextual contingencies of their own beliefs and values (76%), even of their religion (45%) and values (26%), and in many other ways thoughtfully articulated their own uncertainties. Are American adolescents more open-minded than we have given them credit for?

The particular format of the questions may have been key in eliciting this sophistication. The interview protocol, in particular, invited participants to contemplate questions they are not often asked to consider. The framing—for instance, asking participants to think about how to teach science or to imagine having been raised in a different family—may have been familiar enough to keep the teens comfortable, but abstract enough to engage them in deeper metacognition than is their wont. The context of the AOT interview section, at the end of half an hour of telling the interviewer about their lives and desires, may have made the adolescents feel safe enough to be open to alternatives and admit their uncertainties. Yet participants were not clearly led by interviewers into any the key codes discussed above; they could easily have gone through without demonstrating any openness to dissent or epistemic empathy, and indeed, some did. If the interview protocol used in the present study was the source of the greater sophistication exhibited by our participants, then other protocols may be too conservative to yield true estimations of adolescents' capacities.

Perhaps American teenagers in 2015 are particularly pluralistic and aware of the contingency of their beliefs (especially religion) because of our extraordinarily pluralistic, heterogeneous, multicultural society. A genuine shift towards pluralism and evaluativism in only

25 years would be shocking. Yet it has been an eventful quarter century, with marked evidence of increasing pluralism in the form of tolerance. Acceptance of homosexual marriage has doubled in less than twenty years, shooting up from 27% in 1996 to 60% in 2015 (Gallup, 2015), Acceptance of interracial marriage has shot up, too: from 48% in 1992 to 87% in 2014, up from only 4% in 1958 (Gallup, 2014).

The emphasis on critical thinking in the Common Core curriculum may be partly a result of this trend, and, if properly applied, it could accelerate the trend further. However, it is also possible that schools suppress some manifestations of actively open-minded thinking. The best curricula now emphasize critical thinking, which involves drawing on multiple lines of evidence (as demonstrated in the Common Core, both ELA Shifts and Math Shifts). Yet successful critical thinking in middle and high school is often operationalized as pulling out the evidence that supports one's preferred thesis, not considering the evidence on both sides.

For instance, in one relatively high quality Common Core curriculum, middle schoolers are asked to write short essays with either a pro-drone or anti-drone argument, drawing all their evidence from a few pages of text given in class. Consideration of evidence from both sides, or even a nuanced integration of pro-drone and anti-drone arguments, would not fulfill the assignment. Engaging young students in big, complicated issues is a great opportunity for them to practice actively open-minded thinking. Demanding that they take a side on an issue when they have barely scratched the surface of knowledge encourages the opposite. The present study suggests that most adolescents are happy to explore ranges of evidence without pretending they are experts. We should encourage this.

In the current political climate of fierce polarization, the importance of AOT cannot be underestimated. The research presented here offers a little hope for the future. It looks as if many

of our young do value openness to alternative perspectives and have the capacity to seriously consider other views. We must strive to support and encourage them in these endeavors. For without the capacity to listen to and understand those who think otherwise, our democracy cannot function.

Bibliography:

Baron, J. (1988). *Thinking and Deciding, 1st Ed.* Cambridge University Press.

Baron, J. (1993). Why teach thinking? *Applied Psychology: An International Review, 42,* 191-237.

Baron, J., Scott, S., Fincher, K., & Metz, S. E. (2014). Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition, 4* (3), 265-284.

Berzonsky, M.D., Cieciuch, J., Duriez, B., & Soenens, B. (2011) The how and what of identity formation: Associations between identity styles and value orientations. *Personality and Individual Differences, 50,* 295-299.

Berzonsky, M. D., & Neimeyer, G. J. (1988). Identity status and personal construct systems. *Journal of Adolescence, 11*(3), 195-204.

Brown, A. L. (1997). Transforming schools into communities of thinking and learning about serious matters. *American psychologist, 52*(4), 399-413.

Cacioppo, J.T. & Petty, R.E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42, 1,* 116-131.

Cacioppo, J.T.; Petty, R.E.; Feinstein, J. A.; & Jarvis, W.B.G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119, 2,* 197-253.

Cacioppo, J. T.; Petty, R. E.; & Kao, C. F. (1986) The efficient assessment of need for cognition. *Journal of Personality Assessment, 48, 3,* 306-307.

Cazden, C. B. (2001). The language of teaching and learning. *The language of teaching and learning*. Portsmouth, NH: Heinemann.

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. *Advances in experimental social psychology, 24*(20), 1-243.

Cohen, J. A. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37-46.

Common Core State Standards Initiative (2016). About the standards. retrieved from http://www.corestandards.org/about-the-standards/ on 3/14/2016.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalogue of Selected Documents in Psychology, 10,* 85-104.

Day, M. (Sept 11, 2013). Pope Francis assures atheists: You don't have to believe in God to go to heaven. *The Independent.* accessed online 3/5/2016, http://www.independent.co.uk/news/world/europe/pope-francis-assures-atheists-you-don-t-have-to-believe-in-god-to-go-to-heaven-8810062.html

De Grada, E., Kruglanski, A. W., Mannetti, L., & Pierro, A. (1999). Motivated cognition and group interaction: Need for closure affects the contents and processes of collective negotiations. *Journal of Experimental Social Psychology, 35*(4), 346-365.

Dewey, J. (1916). *Democracy and Education: An Introduction to the Philosophy of Education.* New York: Macmillan.

Dijksterhuis, A. P., Van Knippenberg, A. D., Kruglanski, A. W., & Schaper, C. (1996). Motivated social cognition: Need for closure effects on memory and judgment. *Journal of Experimental Social Psychology, 32*(3), 254-270.

Duckworth, A. L. & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality, 45 (3),* 259-268.

Duckworth, A. L. & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Research, 44(4),* 237-251.

Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin, 94*(1), 100.

Erikson, E. (1968). *Identity: Youth and crisis.* New York: Norton.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions.* (2nd ed.) New York: John Wiley.

Foster, G., & Ysseldyke, J. (1976). Expectancy and halo effects as a result of artificially induced teacher bias. *Contemporary Educational Psychology, 1*(1), 37-45.

Gallup (2013). Americans say U.S. schools should teach 'soft' skills. August 21, 2013, by Lopez, S. J. & Calderon, V. J. Retrieved from: http://www.gallup.com/poll/164060/americans-say-schools-teach-soft-skills.aspx

Gallup (2014). In U. S., 87% approve of black-white marriage, vs. 4% in 1958. July 25, 2013, Newport, F. Retrieved 4/1/2015 from http://www.gallup.com/poll/163697/approve-marriage-blacks-whites.aspx

Gallup (2015). Record-high 60% of Americans support same-sex marriage. May 19, 2015 by McCarthy, J. Retrieved 4/1/2015 from http://www.gallup.com/poll/183272/record-high-americans-support-sex-marriage.aspx

Ginossar, Z. & Trope, Y. (1987). Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology, 52,* 464-474.

Godfrey-Smith, P. & Yegnashankaran, K. (2011). Reasoning as deliberative in function but dialogic in structure and origin: Commentary on Mercier & Sperber's "Why do humans reason? Arguments for an argumentative theory." *Behavioral and Brain Sciences, 34,* 80.

Goos, M. (2004). Learning mathematics in a classroom community of inquiry. *Journal for Research in Mathematics Education, 35, 4,* 258-291.

Grotevant, H. D., & Von Korff, L. (2011). Adoptive identity. In *Handbook of identity theory and research* (pp. 585-601). Springer New York.

Heine, S. J.; Lehman, D. R.; Peng, K.; & Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales? The Reference-group effect. *Journal of Personality and Social Psychology, 82,* 6, 903-918.

Hofer, B. & Pintrich, P. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67(1),* 88-140.

Kassarjian, H. H. & Cohen, J. B. (1965). Cognitive dissonance and consumer behavior. *California Management Review, 8,* 55-64.

King, P. M., & Kitchener, K. S. (Eds.). (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults.* San Francisco: Jossey-Bass.

King, P. M., & Kitchener, K. S. (2002). The reflective judgment model: Twenty years of epistemic cognition. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 37–61). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

King, P. M., Kitchener, K. S., & Wood, P. K. (1994). Research on the reflective judgment model. *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults.* San Francisco: Jossey-Bass.

Kitchener, K. S. (1983). Cognition, metacognition and epistemic cognition: A three-level model of cognitive processing. *Human Development, 4,* 222–232.

Kokis, J. V., Macpherson, R., Toplak, M. E., West, R. F., & Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology*, *83*(1), 26-52.

Krettenauer, T. (2004). The role of epistemic cognition in adolescent identity formation: Further evidence. *Journal of Youth and Adolescence, 34, 3,* 185-198.

Kruglanski, A. W., Shah, J. Y., Pierro, A., & Mannetti, L. (2002). When similarity breeds content: need for closure and the allure of homogeneous and self-resembling groups. *Journal of personality and social psychology*,*83*(3), 648.

Kuhn, D. (1991). *The Skills of Argument.* New York: Cambridge University Press.

Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher, 28, 2,* 16-26, 46.

Kuhn, D. (2001). How do people know? *Psychological Science, 12, 1,* 1-8.

Kuhn, D.; Cheney, R; & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development, 15,* 309-328.

Kuhn, D. & Moore, W. (2015). Argumentation as core curriculum. *Learning: Research and Practice, 1:1,* 66-78. DOI: 10.1080/23735082.2015.994254

Kuhn, D.; Zillmer, N.; Crowell, A.; & Zavala, J. (2013). Developing norms of argumentation: Metacognitive, epistemological, and social dimensions of developing argumentive competence. *Learning & Instruction. 31 (4),* 1-41. DOI: 10.1080/07370008.2013.830618

Kunda, Z. (1987). Motivation and inference: Self-serving generation and evaluation of evidence. *Journal of Personality and Social Psychology, 53,* 636-647.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin,108(3),* 480-498.

Landis, J. R., & Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, 363-374.

Lirgg, C. D. (1991). Gender differences in self-confidence in physical activity: A meta-analysis of recent studies. *Journal of Sport and Exercise Psychology*, *13*(3), 294-310.

Marcia, J. E. (1966). Development and validation of ego-identity status. *Journal of personality and social psychology*, *3*(5), 551.

Perry, W (1970). *Forms of Intellectual and Ethical Development in the College Years.* Holt, Rinehart, & Winston, Inc.

Rampey, B.; Dion, G.S.; & Donahue, P.L. (2009). The nation's report card: Trends in academic progress in reading and mathematics 2008. *National Assessment of Educational Progress (NAEP)*, National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

Rokeach, M. (1960) *The Open and Closed Mind.* Basic Books: New York, NY.

Runco, M. (2010) "Divergent thinking, creativity, and ideation." in *The Cambridge Handbook of Creativity*, Kaufman, J. C. & Sternberg, R.J. (Eds.) Cambridge University Press: New York, NY.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54,* 2, 93-105.

Schwartz, S. J., Beyers, W., Luyckx, K., Soenens, B., Zamboanga, B. L., Forthun, L. F., Hardy, S. A., Vazsonyi, A. T., Ham, L. S., Kim, S. Y., Whitbourne, S. K., & Waterman, A. S. (2011). Examining the light and dark sides of emerging adults' identity: A study of identity status differences in positive and negative psychosocial functioning. *Journal of Youth and Adolescence, 40,* 839-859.

Schwartz, S. J., Zamboanga, B. L., Luyckx, K., Meca, A., & Ritchie, R. A. (2013). Identity in emerging adulthood reviewing the field and looking forward. *Emerging Adulthood, 1*(2), 96-113.

Schenkel, S., & Marcia, J. E. (1972). Attitudes toward premarital intercourse in determining ego identity status in college women1. *Journal of Personality*,*40*(3), 472-482.

Schommer, M. (1990). Beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology, 82, 3,* 498-504.

Sisk, C. L., & Zehr, J. L. (2005). Pubertal hormones organize the adolescent brain and behavior. *Frontiers in neuroendocrinology*, *26*(3), 163-174.

Slugoski, B. R., Marcia, J. E., & Koopman, R. F. (1984). Cognitive and social interactional characteristics of ego identity statuses in college males. *Journal of personality and social psychology*, *47*(3), 646.

Stanovich, K. E. & West, R. F. (1997) Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89, 2,* 342-357.

Stanovich, K. E. & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology, 127, 2*, 161-188.

Stanovich, K. E. & West, R. F. (2007) Natural myside bias is independent of cognitive ability. *Thinking and Reasoning, 13 (3)*, 225-247.

Stanovich, K. E. & West, R. F. (2008). On the failure of cognitive ability to predict myside and one-sided thinking biases. *Thinking and Reasoning*, *14, 2*, 129-167.

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*,*131*(1), 159-171.

Tishman, S., Jay, E., & Perkins, D. N. (1993). Teaching thinking dispositions: From transmission to enculturation. *Theory into practice*, *32*(3), 147-153.

Vygotsky, L. (1986). *Thought and language.* Trans. A. Kozulin. MIT Press (Original work published 1934).

Wason, P. C. & Evans, J. St. B. T. (1975). Dual processes in reasoning? *Cognition, 3*, 141-154.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of personality and social psychology*, *67*(6), 1049.

Weinstock, M. (1999). *Epistemological understanding and argumentative competence as foundations of juror reasoning skill.* Unpublished doctoral dissertation. Teachers College, Columbia University, NY. As cited in Kuhn, 2001.

Wenz-Gross, M., Siperstein, G. N., Untch, A. S., & Widaman, K. F. (1997). Stress, social support, and adjustment of adolescents in middle school. *The Journal of Early Adolescence, 17*(2), 129-151.

West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology, 100*(4), 930.

Wyer, R. S. & Frey, D. (1983). The effects of feedback about self and others on the recall and judgments of feedback-relevant information. *Journal of Experimental Social Psychology, 19,* 540-559.

Yegnashankaran, K. (2010). Reasoning as action. Unpublished doctoral dissertation. Harvard University.

APPENDICES

Chapter 1

Evolution Interview Protocol

i.
How did humans come to exist? [*How did that happen?]*

How do you know that?

How sure are you? Is there anything that would make you more sure? Is there anything that would make you less sure? *[Anything else?]*

Do you think most people agree with your own point of view about how humans came to exist, or disagree with it?

Why/Why not?

ii.
OK. Now I want to ask you another question. How did *animals* come to exist? *[If this is the same as how humans came to exist, skip to* ***III:*** *if, different, continue,*] Why is it different from the way humans came to exist?

How do you know that?

How sure are you?

Is there anything that would make you more sure?

Is there anything that would make you less sure? [*Anything else?]*

Do you think most people agree with your own point of view on this issue or disagree with it? [*Why/Why not?]*

Is your explanation for animals also true for trees?

iii.
What is the theory of evolution? *[Prompt generously! E.g. How would that work? Could you explain what you mean by X (where X is a word/phrase interviewee has just used]? Can you say more about that? If they mention "natural selection," ask them to explain what they mean by natural selection.]*

What does it mean that it's a "theory"?

Why do you think a lot of scientists say that evolution is the right explanation for how all of the species came to be?

Is that reasonable of the scientists, or not?

What is the best way to get knowledge about the origins of life? Why?

If you got knowledge that way, could you ever be wrong? Why or why not?

Those are all the questions that I have. Is there anything else that you would like to talk about?

So I asked you these questions because my team is interested in how people around the country think about the origins of life and about evolution. Is there anything else you'd like to add about your view of the origins of life?

Study 2 Criteria for Belief Measure

Which of the following do you think are **acceptable reasons** to believe something is **true**? *Check all that apply*. [Order randomized]

- A teacher told me it is true.
- The Bible says it is true.
- My parents told me it is true.
- It seems reasonable to me.
- It explains a whole lot of things.
- I feel it is true in my heart.
- I have read about it in a book.
- I saw it myself.
- My friends told me they saw it themselves.
- There is good scientific evidence for it.
- A scientist said it is true.
- There is scientific consensus that it is true.
- I have always believed it is true.
- My clergy/minister/priest said it is true.
- Belief that it is true encourages ethical behavior.
- I learned about it from media.

2. Which of the following do you think are **really excellent** reasons to believe something? That is, which of the following do you think could **justify great confidence** that something is true?

Check at most the BEST **THREE.**

(same responses as above)

Study 3 Criteria for Belief Measure

Below are some possible reasons why someone might believe something. For each, choose whether it would be a good reason to believe something is true, or not a good reason.

1. If *it feels true in your heart,* that is a(n):

2. If *there is good scientific evidence for it*, that is a(n):
Response scale: Excellent reason for belief, Good reason for belief, Okay reason for belief, Bad reason for belief.

Study 4 Criteria for Belief Measure

Below are 20 possible reasons you might believe something and **four boxes.** For each, decide whether it is an **excellent reason** to believe that something is true, a **good reason,** an **okay reason**, or a **bad reason**. *Drag and drop* each item into the appropriate box (there is one box for each category).

ITEMS *(Order randomized for participants)*:

Religious Criteria:

1. My religious leader said it is true.
2. The Bible says it is true.

Scientific Criteria:

3. There is scientific consensus that it is true.
4. There is good scientific evidence for it.
5. Scientists agree that it is true.

Heart Criteria:

6. It feels true in my heart.
7. My heart tells me it is true.

Affiliative Criteria:

8. My parents told me it is true.
9. The people I love believe it is true.
10. If I didn't believe it, my friends and family would be upset.
11. I don't know anyone who doesn't believe it.

Explanatory Power Criteria:

12. It explains a whole lot of things.
13. It explains a lot of the things I've seen.
14. It is the simplest explanation I know for some things.

Other Criteria:

15. Belief that it is true would encourage people to be good.
16. It fits with my other beliefs.
17. I have always believed it is true.
18. It makes sense to me.
19. I saw it happen.
20. I read it in the newspaper.

Chapter 3

Adolescent Actively Open-Minded Thinking Scale (AAOT Scale)

1. I can learn things by listening to people I disagree with.
2. People should pay attention to new possibilities.
3. When I disagree with someone, I try to understand why they think as they do.
4. I enjoy conversations with people with different beliefs about the way the world works.
5. I would like to talk to people from other countries about what it's like where they come from.
6. I respect people who change their mind because of a good argument.
7. When someone disagrees with me, I have less respect for that person. (reverse scored)
8. Changing your mind is a sign of weakness. (reverse scored)